

Mon Jun 26 09:10:26 2000

us-09-214-478-1.rge

Shin-din Chen

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 23, 2000, 06:06:58 ; Search time 878.41 Seconds

(without alignments)
-980.090 Million cell updates/sec

Title: US-09-214-478-1

Perfect score: 885

Sequence: 1 atgactcgtccggtcc.....atgactcaccatgtag 885

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 1765538

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

GenEmbl.*
1: gb_ba1.*
2: gb_ba2.*
3: gb_om.*
4: gb_ov.*
5: gb_pat.*
6: gb_ph.*
7: gb_pl1.*
8: gb_pl2.*
9: gb_pr1.*
10: gb_pr2.*
11: gb_pr3.*
12: gb_ro.*
13: gb_sts.*
14: gb_sy.*
15: gb_un.*
16: gb_vi.*
17: em_fun.*
18: em_hum1.*
19: em_hum2.*
20: em_in.*
21: em_om.*
22: em_or.*
23: em_ov.*
24: em_pat.*
25: em_ph.*
26: em_pl.*
27: em_to.*
28: em_sts.*
29: em_sy.*
30: em_un.*
31: em_vi.*
32: gb_htg1.*
33: gb_htg2.*
34: gb_in1.*
35: gb_in2.*
36: em_ba1.*
37: em_ba2.*
38: em_hum3.*
39: em_hum4.*
40: gb_pr4.*
41: gb_htg3.*
42: gb_htg4.*
43: gb_htg5.*
44: gb_htg6.*

45: gb_htg7.*
46: em_htg1.*
47: em_htg2.*
48: em_htg3.*
49: em_hum5.*
50: gb_pl3.*
51: gb_pr5.*
52: gb_htg8.*
53: gb_htg9.*
54: gb_htg10.*
55: gb_htg11.*
56: gb_htg12.*
57: gb_htg13.*
58: gb_htg14.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	885	100.0	3189	5 A52460	A52460 Sequence 4
C 2	885	100.0	8299	5 AR009152	AR009152 Sequence
C 3	885	100.0	35935	16 ADRCOMPEN	M73260 Mastadenovi
C 4	875.4	98.9	35937	16 ADRCG	J01917 Adenovirus
C 5	328.8	37.2	3119	16 S82508	S82508 9E4: orf2..
C 6	319.2	36.1	35100	16 AF108105	AF108105 Human ade
C 7	275.4	31.1	34125	16 AT12CGA	Y73487 Adenovirus
C 8	273.8	30.9	3498	16 ADER24	X51800 Adenovirus
C 9	222.2	25.1	34214	16 ADRGENOME	L19443 Human adeno
C 10	65	7.3	34094	16 AB026117	AB026117 porcine a
C 11	65	7.3	34094	16 PAD237815	AJ237815 Porcine a
C 12	62.2	7.0	3028	31 PRORR8G	L43363 Mastadenovi
C 13	62.2	7.0	3028	31 PRORR8G	AF083132 Porcine a
C 14	49.8	5.6	31323	16 CAU77082	U77082 Canine aden
C 15	48.4	5.5	30288	16 CAU55001	U55001 Canine aden
C 16	48.4	5.5	30536	16 CAV1GENOM	Y07760 Canine aden
C 17	37.8	4.3	34446	16 AF030154	AF030154 Bovine ad
C 18	36.4	4.1	10499	2 AE000423	AE000423 Escherich
C 19	36.4	4.1	74377	43 AC016797	AC016797 Homo sapi
C 20	36.4	4.1	74377	43 AC016797	AC016797 Homo sapi
C 21	36.2	4.1	55123	42 AC016343	AC016343 Homo sapi
C 22	36	4.1	3181	34 AB023583	AB023583 Drosophill
C 23	36	4.1	3654	35 AF125986	AF125986 Drosophill
C 24	36	4.1	9282	35 AF083334	AF083334 Antherea
C 25	36	4.1	12644	42 AC014231	AC014231 Drosophill
C 26	36	4.1	136360	32 AL136360	AL136360 Homo sapi
C 27	36	4.1	149886	12 MMHC438N12	AF049850 Mus muscu
C 28	36	4.1	153442	54 AC015838	AC015838 Homo sapi
C 29	35.4	4.0	123937	42 AC008145	AC008145 trypanoso
C 30	35.2	4.0	155083	41 AC007847	AC007847 Homo sapi
C 31	35	4.0	180009	44 AC011777	AC011777 Homo sapi
C 32	35	4.0	234131	54 AC022389	AC022389 Homo sapi
C 33	34.8	3.9	2041	16 HS1732	M1316 HSV-1 (KOS)
C 34	34	3.8	468	3 BOVAMEL	J02695 Bovine amel
C 35	34	3.8	53793	40 AC007875	AC007875 Homo sapi
C 36	34	3.8	133952	32 AL135905	AL135905 Homo sapi
C 37	34	3.8	309026	41 AC008760	AC008760 Caenorhab
C 38	33.8	3.8	1654	16 MEAPHOSPPA	M89920 Measles vir
C 39	33.8	3.8	9207	16 MEANIBE	X16567 Measles vir
C 40	33.8	3.8	9208	16 MEANIBE	X16569 Measles v
C 41	33.8	3.8	15894	16 AB012948	AB012948 Measles v
C 42	33.8	3.8	15894	16 AB012949	AB012949 Measles v
C 43	33.8	3.8	184762	57 AC016946	AC016946 Homo sapi
C 44	33.8	3.8	221563	52 AC010746	AC010746 Homo sapi
C 45	33.6	3.8	149093	45 AC019293	AC019293 Homo sapi

ALIGNMENTS

[illegible]

ETISVSALGQEDTGSLEATLNYLTNRROKIPSLHSNSEERILRTVQOVSYNLNL
RGVTPSVALDMTARMPCMYASNPFRINRLMDYLHRAAAYNPFTNAILNPHWLP
PGFTYTGGEVPEGNDGLWDDSDVSFQSPQPTLLEQOQREAAALRKESFRPSS
LSLGAAPRSPDASSPFLSLIGSLSTRTTRPRLLEGEYLNLSLQPOREKRLPPAF
PNNIGSLVDKMSRWKIYAQERHDYVPGPPPTTRQRHQRGLVWEDDDSDADDSSVLD
LGGSGNPFALRPLGRMF
16286..16474
/codon_start=1
/product="protein pVII"
/protein_id="AAA96408.1"
/db_xref="GI:209845"
/translation="MRAARLAAGIVTPPRSRRAAAAAAIAISANTQRRGNVTWY
RDSVSGLRVPVTRPPRN"
16544..17650
/codon_start=1
/product="protein v"
/protein_id="AAA96409.1"
/db_xref="GI:209846"
/translation="MSRKIKEMQVIAPEIYIGPKKEQDYKPKRLKRVKKKKDD
DELDDEVELLHATAPRRRVQWGRVRLRPGTTVFTGGERSTRTKRVYDEVY
DELDLEQANERLGEFAYGRKHKMLAPLDEGNPTSLKPVTLQOVLPAALPEEKRG
LKRESGLAFTVQLMVKRQRLDEKMTVEKLEPEVAVRIRIKVAPGLGVQTVDV
QIPTSTSIATATEMETQTSVAVADAQAQAAAAASKTSTEVTDPMFVRYSA
RRRGRSKYGAASALLPEVALHPSIAPTPYRGYTYRPRRRATRRRTTTGTGRRRR
RRQPVLAISVRYVAREGGRTLVLPRTARYHPSIV"
17774..17920
/codon_start=1
/product="protein 5 precursor"
/protein_id="AAA96410.1"
/db_xref="GI:209847"
/translation="MRAHRRRRRRSHRRMRGGILPLLIPLIAAAAGVPGIASVALQ
AQRH"
18003..18755
/codon_start=1
/product="protein pVI"
/protein_id="AAA96411.1"
/db_xref="GI:209848"
/translation="MEDINFASLAPRHGRPFMGWQDICTGSMGSAFSGSLWSGI
KNEFSVKNVSKANSTGOMLRDLKAEQNFQKQVVDGLASGISGVVDLANQVQNK
INSLPRPPEEPVAVETVSPGEGKRRPRDRETLVTQIDEPYSYEAALQGLP
TSLPIAPMAIGVIGQHTVLDLPADTQOKPVLPGPTAVVTVTPSRASLRRASGP
RSLRVASGNWQSTLNSIVGLGVQSLKRRFC"
25819..26484
/note="100 K protein"
/codon_start=1
/protein_id="AAA96412.1"
/db_xref="GI:209849"
/translation="ESCIIGTFELQSPDEKSAAPGLKTPGLWTSAYLRFVPEDY
HAHEIRFYEDQSRPNAELTACVITQGHILGLOALNKARQELLKRGVYLDQPSG
EELNPIPPPPYQOQPRALASQDQTKRAAAATHGGILGSGRGGRGFGGSGH
DGRLEPRGRSGRGRVRRNTVTLGRIPLAGAPEIGNRFRQHGYNLRSSGAAGTARSP
TOP"
26485..26514
/partial
/note="of a 33 K protein"
/number=1
26713..27083
/note="of a 33 K protein"
/number=2
27174..27857
/codon_start=1
/product="protein pVIII"
/protein_id="AAA96413.1"
/db_xref="GI:454806"
/translation="MSKEIPTMYMSYQPMGLAQAQDYSTRINTNSAGPHWISV
NGIRAHNRILLEQAATITPNNLRNPALVYQESPAFTVVLPRDAQAEVQMT
NSGAQAGGFHVRSPCOGITHLTIRGQFNLDNDESSSLGLRQDGTQIGAGRP
SFTPQALITQTSSEPRSGGIGTLQIEEFVSPVYVFNPFSGPGHYPDQIPNEDA
VKDSADGYD"
8367 a 10073 c 9761 g 7734 t

CDS
CDS
CDS
exon
exon
CDS
BASE COUNT
ORIGIN

3822 TTGCCAGCTGGTTAGATGGTGGATGGCCCATGTTTAAATCAGAGGTTATATGG 3881
481 taccggaggtgggtgaattacaacatgccaaagaggaatgtttatgccagcggttt 540
3882 TACCGGGAGGTGGTGAATACACATGCCAAAGAGTAATGTTATGCCAGCGTGT 3941
541 atgaggggtgcccacttaactcactcgttggatgatggccacgtgggtttctgtg 600
3942 ATGAGGGGTGCCACTTAATCTACTCGCTGTGGTGTGATGATGCCACGTGGTCTGTG 4001
601 gtcccccgcattgagcttggatagacagcccttgcaactgtgggattttgacaattattgtg 660
4002 GTCCCGCCCATGAGCTTGGATACAGCGCCCTGACCTGTGGGATTTTGAACAATATTGTG 4061
661 gtgtgtgtcagcttactgtcgtgatttaagtagtagatcaggggtgcgtgctgtgcccgg 720
4062 GTGCTGTCTGCAGTACTGTCTGCTGATTAAGTAGATCAGGGTGCCTGTGTGCCCGG 4121
721 agacaagggccttatctgctggggcgtgctgcaatcatcgtgsggagaccactgcatg 780
4122 AGACAAGGCGCTTATGCTGCGGCGGTGCGAATCATCGCTGAGGACCATGCCATG 4181
781 tigtattctcagacagagggggggggggcagcagttattcgcgcgtcgtcgcagcac 840
4182 TTGTATTCTCGAGGACGAGCGGGGGGCGGACAGTATTATTCGCGCCTGCTGCAGCAC 4241
841 caccgcctactcgtgacacattatgacttacttaccctccatgtag 885
4242 CACCGCCCTATCTGATGACGATGATGACTTACCTACCCCATGTAG 4286

RESULT 3
ADRCOMPEN/c 35935 bp DNA VRL 08-APR-1996
LOCUS Mastadenovirus h5 gene, complete genome.
DEFINITION W73260 M29978
ACCESSION W73260.1 GI:209842
VERSION protein 5 precursor; protein IIIa; protein V; protein pVI; protein
KEYWORDS pVII; protein pVIII.
HUMAN adenovirus 5 DNA.
SOURCE Human adenovirus type 5
ORGANISM Human adenovirus type 5
REFERENCE 1 (bases 1 to 35935)
AUTHORS Chroboczek, J., Bieker, F. and Jacrot, B.
TITLE The sequence of the genome of adenovirus type 5 and its comparison
with the genome of adenovirus type 2
JOURNAL Virology 186 (1), 280-285 (1992)
MEDLINE 92087470
FEATURES Location/Qualifiers
1..35935
/organism="Human adenovirus type 5"
/db_xref="taxon:28285"
11565..12297
/note="52/55 K protein gene"
/codon_start=2
/protein_id="AAA96406.1"
/db_xref="GI:209843"
/translation="SFNNHRTIVAREEVAIGLHMLWDFVSALEQNPNSKPLMAQLFL
IVQSRNEDAFRDLALNIVEPEGWLLDLNLIQSVQERSLSADKVAANISMLS
LGKFKYRIHTPIVPIDKEVKIEGFMRLAKVLTLSDDLGVYRNIHKAVSVSR
RELSDELHMSLQRLADPGTSGDREAESIFDAGADLNRAPSRRALEAGAGPLAVAP
ARAGNVGVEYDEDEYEPEDEGY"
12318..14075
/codon_start=1
/product="protein IIIa"
/protein_id="AAA96407.1"
/db_xref="GI:289094"
/translation="MMQDATDPVRAALQSPGSLNSTDDWRQVMDRIMSLTARNPDA
FRQOPQANRLSAILAEVVPARANTHEKVLAIIVNALAENRAIRDEAGLYDALLQRY
ARYNSGVQTNLDRLGVDRVAOQOQGNLGSVALNAPLSTQPNAPRGQD
YNTFVSALRLAWTETPOSEVYQSGPDYFOTSRQLOTLNLSQKLNQGLWGYRAP
GDRATVSSLLTNSRLLLLLLTPTDSDGSVSRTDYLGHLLTLYREALGQARVDEHTFO

Query Match 100.0%; Score 885; DB 16; Length 35935;
 Best Local Similarity 100.0%; Pred. No. 3e-246;
 Matches 885; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgactacgtccggcgcacattggcagatgacactacgacacacacgacatcgtgtgtt 60
 DB 34077 ATGACTACGTCCGGCGTTCATTGGCATGACACTACGACCAACACGATCTCGGTGTCT 34018

QY 61 cgggcactccgtacagtaggagatgctacacctcttggagacagaaacccgcgtacc 120
 DB 34017 CGGCGCACTCCGTACAGTAGGAGTCTCTACCTCTCTTTTGACAGCAAAACCCGCTACC 33958

QY 121 atactggagatcatcgcgtgctgcgaatgtaacactttgacacattgacaaactgagt 180
 DB 33957 ATACTGGAGATCATCGCTGTGTCGCGAATTAACACTTTGACAATGCACACGTGAGT 33998

QY 181 tacgtgcgaggtttccctgcagtgaggatttaccgattacgaagaatgggttttccc 240
 DB 33897 TAGCTCGAGGTTCCCTGCAGTGTGGATTTACGCTGATTCAGGAATGGTGTGTTCCC 33838

QY 241 tgggatatgtttccctgcagtgaggatttaccgattacgaagaatgggttttccc 300
 DB 33837 TGGGATATGGTTTCAACGCGGAGGAGTGTGAATCCTCAGAGTGTATCAGCTGTGC 33778

QY 301 ctgtgttgccacatgatcatcatgacgacgatgatccatgttaccagtgatccctgg 360
 DB 33777 CTGTGTGTGCGCAACATTGATATCATGACGACATGATCCATGTTACGATCCTGG 33718

QY 361 gatttccactgtcattgttcacgtcccggttccctgcagtgatagcggcgagcagtt 420
 DB 33717 GCTCTCCACTGTCTATGTTCAGTCCCGGTCCCTGCAGTGTATAGCGCGGCGCAGTT 33658

QY 421 ttggccagctggtttaggatggtgtgagtgcccatgttttaacagaggtttatgg 480
 DB 33657 TTGGCCACTGTTAGGATGCTGTGATGCGCCCATGTTTAAATCAGAGTTTATATGG 33598

QY 481 tacggggagggttggaattacaacatgccaaagagtaagtatttatgtccacggttt 540
 DB 33597 TACCGGAGGTGTTGTAATTAACAATGCCAAAGAGTATGTTATGTCCACGGTGT 33538

QY 541 atgaggggtgcacctaactacgtcgcgttggatgatggccacgtgggtttctgtg 600
 DB 33537 ATGAGGGGTGCCACTTAATCTACCTGCTGTGTGTATGATGCCACGTGGTCTGTG 33478

QY 601 gtccccccatgagcttggatagacagcgccttgcactgtggatttgaacaaattgtg 660
 DB 33477 GTCCCCCATGAGCTTGGATACAGCGCTTCACGTGGGATTTGAACAAATTTGTG 33418

QY 661 gtctgtgtgcagttactgtgtgatttaagttagatcaggggtgcgtgtgtgcccg 720
 DB 33417 GTCTGTGTGCAATGTTACTGTGTGATTTAAGTGAATCAGGCTGCTGTGCCCGG 33358

QY 721 aggaagggcccttatgtctgcggcggtgcaaatcatcgtcagagacacactgcctg 780
 DB 33357 AGGCAAGGCGCTTATGTCTGCGGGCGGTGCCAATCATCTGCTGAGAGACCACTGCCATG 33298

QY 781 ttgtattctgagagcagcagcggcgagcaggtttattcgcgcctctcgtcagcac 840
 DB 33297 TTGTATTCTGAGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 33238

QY 841 caccgccttatcctgatgcacattatgacttaccctaccctatgag 885
 DB 33237 CACGCGGCTTCTCTGATGACGATGATGACTCTACCCCGCATGTAG 33193

RESULT 4
 ADRG/c ADRG 35937 bp DNA VRL 14-MAR-1996
 LOCUS Adenovirus type 2, complete genome.
 DEFINITION J01917 J01918 J01919 J01920 J01921 J01922 J01923 J01924 J01925
 ACCESSION J01926 J01927 J01928 J01929 J01930 J01931 J01932 J01933 J01934
 J01935 J01936 J01937 J01938 J01939 J01940 J01941 J01942 J01943
 J01944 J01945 J01946 J01947 J01948 J01949 J01950 J01951 J01952

VERSION
KEYWORDSSOURCE
ORGANISMREFERENCE
AUTHORS

TITLE

JOURNAL
MEDLINEREFERENCE
AUTHORS

TITLE

JOURNAL
MEDLINEREFERENCE
AUTHORS

TITLE

JOURNAL
MEDLINEREFERENCE
AUTHORS

TITLE

JOURNAL
MEDLINEREFERENCE
AUTHORS

TITLE

JOURNAL
MEDLINEREFERENCE
AUTHORS

TITLE

JOURNAL
MEDLINEREFERENCE
AUTHORS

TITLE

JOURNAL
MEDLINEREFERENCE
AUTHORS

TITLE

JOURNAL
MEDLINEREFERENCE
AUTHORS

TITLE

JOURNAL
MEDLINEREFERENCE
AUTHORS

TITLE

J01953 J01954 J01955 J01956 J01957 K00086 K00394 K00395 K02367
 M13004 V00007 V00008 V00009 V00010 V00011 V00012 V00013 V00014
 V00015 V00016 V00017 V00018 V00019 V00020 V00023 V00024
 J01917.1 GI:209811
 DNA polymerase; DNA-binding protein; RNA polymerase III;
 alternative splicing; coat protein; complete genome; genome-linked
 protein; glycoprotein; overlapping genes; polymerase; terminal
 repeat; unidentified reading frame; virus-associated RNA.
 Human adenovirus type 2.

Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
 1 (sites)
 One, K. and Weissman, S.M.
 The nucleotide sequence of a low molecular weight ribonucleic acid
 from cells infected with adenovirus 2

J. Biol. Chem. 246 (22), 6991-7009 (1971)
 72060437
 2 (sites)
 Jornvall, H., Ohlsson, H. and Philipson, L.
 An acetylated N-terminus of adenovirus type 2 hexon protein

Biochem. Biophys. Res. Commun. 56 (2), 304-310 (1974)
 74147480
 3 (sites)
 Celma, M.L., Pan, J. and Weissman, S.M.
 Studies of low molecular weight RNA from cells infected with

adenovirus 2. I. The sequences at the 3' end of VA-RNA I
 J. Biol. Chem. 252 (24), 9032-9042 (1977)
 78046048
 4 (sites)
 Celma, M.L., Pan, J. and Weissman, S.M.
 Studies of low molecular weight RNA from cells infected with

adenovirus 2. II. Heterogeneity at the 5' end of VA-RNA I
 J. Biol. Chem. 252 (24), 9043-9046 (1977)
 78046049
 5 (bases 10514 to 10680)
 Pan, J., Celma, M.L. and Weissman, S.M.
 Studies of low molecular weight RNA from cells infected with

adenovirus 2. III. The sequence of the promoter for VA-RNA I
 J. Biol. Chem. 252 (24), 9047-9054 (1977)
 78046050
 6 (bases 30812 to 30900)
 Zain, B.S. and Roberts, R.J.
 Characterization and sequence analysis of a recombination site in

the hybrid virus Ad2+ND
 J. Mol. Biol. 120 (1), 13-31 (1978)
 78153757
 7 (sites)
 Akusjarvi, G. and Pettersson, U.
 Nucleotide sequence at the junction between the coding region of

the adenovirus 2 hexon messenger RNA and its leader sequence
 Proc. Natl. Acad. Sci. U.S.A. 75 (12), 5822-5826 (1978)
 79096069
 8 (bases 5986 to 6236)
 Ziff, E.B. and Evans, R.M.
 Coincidence of the promoter and capped 5' terminus of RNA from the

adenovirus 2 major late transcription unit
 Cell 15 (4), 1463-1475 (1978)
 79084199
 9 (bases 21607 to 21816)
 Akusjarvi, G. and Pettersson, U.
 Sequence analysis of adenovirus DNA. I. Nucleotide sequence at the

carboxy-terminal end of the gene for adenovirus type 2 hexon
 Virology 91 (2), 477-480 (1978)
 79119384
 10 (bases 1 to 156; 35804 to 35937)
 Arrand, J.R. and Roberts, R.J.
 The nucleotide sequences at the termini of adenovirus-2 DNA

J. Mol. Biol. 128 (4), 577-594 (1979)
 79175344
 11 (sites)
 Zain, S., Sambrook, J., Roberts, R.J., Keller, W., Fried, M. and
 Dunn, A.R.
 Nucleotide sequence analysis of the leader segments in a cloned

JOURNAL MEDLINE REFERENCE AUTHORS TITLE	copy of adenovirus 2 fiber mRNA	JOURNAL MEDLINE REFERENCE AUTHORS TITLE	Nucleotide sequence of the EcoRI D fragment of adenovirus 2 genome
Cell 16 (4), 851-861 (1979)	79211249	Cell 16 (4), 841-850 (1979)	81053687
12 (bases 6039 to 6079; 7101 to 7172; 9634 to 9723; 18802 to 18861)	79211249	13 (sites)	24 (sites)
Akusjarvi, G. and Pettersson, U.	79211249	Shinagawa, M. and Padmanabhan, R.	Akusjarvi, G., Mathews, M.B., Andersson, P., Vennstrom, B. and Pettersson, U.
Sequence analysis of adenovirus DNA: complete nucleotide sequence of the spliced 5' noncoding region of adenovirus 2 hexon messenger RNA	79211249	Nucleotide sequence at the inverted terminal repetition of adenovirus type 2 DNA	Structure of genes for virus-associated RNAI and RNAII of adenovirus type 2
Cell 16 (4), 841-850 (1979)	79211249	Biochem. Biophys. Res. Commun. 87 (3), 671-678 (1979)	Proc. Natl. Acad. Sci. U.S.A. 77 (5), 2424-2428 (1980)
13 (sites)	79211248	79209154	80234635
Shinagawa, M. and Padmanabhan, R.	79211248	14 (bases 25634 to 27376)	25 (sites)
Nucleotide sequence of the inverted terminal repetition of adenovirus type 2 DNA	79211248	Galibert, F., Herisse, J. and Courtols, G.	Perricaudet, M., Le Moulllec, J.M. and Pettersson, U.
79209154	79211248	Nucleotide sequence of the EcoRI-F fragment of adenovirus 2 genome	Predicted structure of two adenovirus tumor antigens
15 (sites)	80004828	Gene 6 (1), 1-22 (1979)	Proc. Natl. Acad. Sci. U.S.A. 77 (7), 3778-3782 (1980)
80004828	80004828	15 (sites)	81054654
Sequences from the beginning of the fiber messenger RNA of adenovirus-2	80004828	15 (sites)	26 (sites)
J. Mol. Biol. 131 (2), 341-352 (1979)	80004828	Zain, B.S. and Roberts, R.J.	Anderson, C.W. and Lewis, J.B.
80039717	80004828	Sequences from the beginning of the fiber messenger RNA of adenovirus-2	Amino-terminal sequence of adenovirus type 2 proteins: hexon, fiber, component IX, and early protein 1B-15K
16 (bases 5909 to 6178; 7023 to 7212; 9452 to 9836)	80039717	17 (sites)	Virology 104 (1), 27-41 (1980)
Akusjarvi, G. and Pettersson, U.	80039717	Structure of two spliced mRNAs from the transforming region of human subgroup C adenoviruses	Cell 22 (1 PT 1), 157-170 (1980)
Sequence analysis of adenovirus DNA. IV. The genomic sequences encoding the common tripartite leader of late adenovirus messenger RNA	80039717	18 (sites)	81042299
J. Mol. Biol. 134 (1), 143-158 (1979)	80039717	18 (sites)	28 (sites)
80140449	80039717	Perricaudet, M., Akusjarvi, G., Virtanen, A. and Pettersson, U.	Persson, H., Jorvall, H. and Zabielski, J.
Structure of two spliced mRNAs from the transforming region of human subgroup C adenoviruses	80039717	Nature 281 (5733), 694-696 (1979)	Multiple mRNA species for the precursor to an adenovirus-encoded glycoprotein: identification and structure of the signal sequence encoding the common tripartite leader of late adenovirus messenger RNA
80045037	80039717	19 (sites)	Proc. Natl. Acad. Sci. U.S.A. 77 (11), 6349-6353 (1980)
Cell 18 (2), 569-580 (1979)	80045037	19 (sites)	81101049
80045037	80045037	Baker, C.C., Herisse, J., Courtols, G., Galibert, F. and Ziff, E.	Butner, W. and Veres-Molnar, Z.
Messenger RNA for the Ad2 DNA binding protein: DNA sequences encoding the 5' leader and heterogeneity at the mRNA 5' end	80045037	19 (sites)	29 (sites)
Cell 18 (2), 569-580 (1979)	80045037	20 (bases 3504 to 4109)	Butner, W. and Veres-Molnar, Z.
80045037	80045037	20 (bases 3504 to 4109)	Localization of the 3'-terminal end of the EcoRI B fragment-specific early mRNA of adenovirus type 2
20 (bases 3504 to 4109)	80045037	20 (bases 3504 to 4109)	FEBS Lett. 122 (2), 317-321 (1980)
80045037	80045037	20 (bases 3504 to 4109)	81138857
Determination and analysis of adenovirus-2 DNA sequences which may include signals for late messenger RNA processing	80045037	20 (bases 3504 to 4109)	30 (sites)
J. Mol. Biol. 135 (2), 413-433 (1979)	80045037	20 (bases 3504 to 4109)	Akusjarvi, G., Zabielski, J., Perricaudet, M. and Pettersson, U.
80045037	80045037	20 (bases 3504 to 4109)	The sequence of the 3' non-coding region of the hexon mRNA discloses a novel adenovirus gene
20 (bases 3504 to 4109)	80045037	20 (bases 3504 to 4109)	Nucleic Acids Res. 9 (1), 1-17 (1981)
80045037	80045037	20 (bases 3504 to 4109)	81150446
20 (bases 3504 to 4109)	80045037	20 (bases 3504 to 4109)	31 (bases 459 to 608)
80045037	80045037	20 (bases 3504 to 4109)	Osborne, T.F., Schell, R.E., Burch-Jaffe, E., Berget, S.J. and
20 (bases 3504 to 4109)	80045037	20 (bases 3504 to 4109)	Query Match
80045037	80045037	20 (bases 3504 to 4109)	Best Local Similarity
20 (bases 3504 to 4109)	80045037	20 (bases 3504 to 4109)	Matches
80045037	80045037	20 (bases 3504 to 4109)	879; Conservative
20 (bases 3504 to 4109)	80045037	20 (bases 3504 to 4109)	Score 875.4; DB 16; Length 35937;
80045037	80045037	20 (bases 3504 to 4109)	Pred. No. 1.9e-243;
20 (bases 3504 to 4109)	80045037	20 (bases 3504 to 4109)	Mismatches 6; Indels 0; Gaps
80045037	80045037	20 (bases 3504 to 4109)	34018
20 (bases 3504 to 4109)	80045037	20 (bases 3504 to 4109)	34018
80045037	80045037	20 (bases 3504 to 4109)	34018
20 (bases 3504 to 4109)	80045037	20 (bases 3504 to 4109)	34018
80045037	80045037	20 (bases 3504 to 4109)	34018
20 (bases 3504 to 4109)	80045037	20 (bases 3504 to 4109)	34018
80045037	80045037	20 (bases 3504 to 4109)	34018

```
QY 301 ctgtgtgtgtgccaacattgatataatgacagagcatgatccatgcttaccagctctgg 360
D 33777 CTGTGTGTGTGCAACATGATATCATGACGAGCATGATCATGCTTACGAGTCTGTG 33718
QY 361 gctctcaactgcatgttccacgtcccggtccctcagtgatagccgagcaggtt 420
D 33717 GCTCTCCACTGTCATGTTCTTCAGTCCCGGTCCCTCAGTCAGTACAGCCGGCAGGTT 33658
QY 421 ttggccagctggtttaggatggtggtggtggtggtggtggtggtggtggtt 480
D 33657 TTGGCCAGCTGTTTAGGATGTTGTTGATGTTGATGTTGATGTTGATGTTGATGTT 33598
QY 481 taccggaggtgctgaattacaacatgccaagaaggttaattgatttccagcgtgttt 540
D 33597 TACCGGAGGTGTTGAATTACAACTATGATGATGATGATGATGATGATGATGATGATG 33538
QY 541 atgaggggtgcacacttaatactacgtggtggtggtggtggtggtggtggtt 600
D 33537 ATGAGGGGTGCGCACTTAATCTACCTGCGCTGTGTATGATGATGATGATGATGATG 33478
QY 601 gtcccccgcacatgagcttggatgacacagccctgacgtggtggtggtggtggtt 660
D 33477 GTCCCGGCATGAGCTTGGATGATGATGATGATGATGATGATGATGATGATGATG 33418
QY 661 gtctgtgtgctgactgctgctgctgctgctgctgctgctgctgctgctgctg 720
D 33417 GTCTGTGTGCTGAGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 33358
QY 721 aggaacagccttctgctgctgctgctgctgctgctgctgctgctgctgctg 780
D 33357 AGGACAGCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 33298
QY 781 ttgtattctgacagcagcagcagcagcagcagcagcagcagcagcagcagcagc 840
D 33297 TTGTATTCTGACAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGC 33238
QY 841 caccgctctctgacgacgacgacgacgacgacgacgacgacgacgacgacgac 885
D 33237 CACCGCTCTCTGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 33193

RESULT 5
S82508
LOCUS S82508 3119 bp DNA VRL
DEFINITION 9E4: orf2. . orf7 [adenovirus type 9, Genomic, 6 genes, 3119 nt].
ACCESSION S82508
VERSION S82508.1 GI:1699390
KEYWORDS
SOURCE Human adenovirus type 9.
ORGANISM Human adenovirus type 9.
REFERENCE 1 (bases 1 to 3119).
AUTHORS Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
TITLE Mammory tumors induced by human adenovirus type 9: a role for the
JOURNAL Viral early region 4 gene
MEDLINE Breast Cancer Res. Treat. 39 (1), 57-67 (1996)
REMARK GenBank staff at the National Library of Medicine created this
entry [NCBI gibbsq 178541] from the original journal article.
This sequence comes from Fig. 1.
Region: 9E4.
Location/Qualifiers
1. .3119
/organism="Human adenovirus type 9"
/db_xref="taxon:10527"
471. .848
/gene="E4"
/notes="early region 4"
471. .848
/gene="E4"
/notes="orf1 transforming protein; This sequence comes from
Fig. 1"
/codon_start=1

QY 46 cgatctcgtgtgtctcgcacacgacagaggtgacgctacacgtctcttcttctgagaca 105
```

```
/protein_id="AAB37504.1"
/db_xref="GI:1699391"
/translation="MAESLYAFIDSPGAIAPVQEGTNRXTFFCPESFHIPPHGVLV
HLKVSIVPTGYQGRFALNDYHARDILITQSDVIFAGRRQELTVLLFNHDFLYVHK
GHPVGTLLERLVIFPSPKIALV"
889. .1281
/gene="orf2"
889. .1281
/gene="orf2"
/notes="This sequence comes from Fig. 1"
/codon_start=1
/protein_id="AAB37505.1"
/db_xref="GI:1699392"
/translation="MLORRGVSYHIVPGVLVYLEDFTOMIKELPRFTHILEG
ITGTRKAYSSMOFLGASFGALAFSLTASPTLSPGSELSAVVAQDLSDFLOITLRE
LRAGRTLLNLVNLTLQVVEQDLELL"
1278. .1631
/gene="orf3"
1278. .1631
/gene="orf3"
/notes="This sequence comes from Fig. 1"
/codon_start=1
/protein_id="AAB37506.1"
/db_xref="GI:1699393"
/translation="MKVCLIMKVEGALWELFMCGVDLHQFVEIIQGMKNENYLMV
QECNMLDEIDGGPAFNVILMDVRVEPLEATVEHLENRVGFDLAVCFHQHSGERL
1634. .1996
/gene="orf4"
1634. .1996
/gene="orf4"
/notes="This sequence comes from Fig. 1"
/codon_start=1
/protein_id="AAB37507.1"
/db_xref="GI:1699394"
/translation="MVLPLPPPLNDRQGSINMGWYRVLADVMGIRMDGFFISS
DAEILLNLREWMYFSWTERQQRKDRRGICCSRAHFCWQKYKRIHYNARD
SIOLAPSSISOGPFTTI"
1926. .2804
/gene="orf6"
1926. .2804
/gene="orf6"
/notes="This sequence comes from Fig. 1; protein sequence
is in conflict with the conceptual translation;
mismatches [125[Q->R],126[T->A],132[S->A],147[H->R]]"
/codon_start=1
/protein_id="AAB37508.1"
/db_xref="GI:1699395"
/translation="MOTEIQSSILRHHPYRRARLPRSDDETRASLTQHPILLPCDHA
DYHVSVRGLPCAAGFTLLQEPFVPMWMLTPEELIKKRCMSVCLCPATLDLVRAQ
MVGYSERWILHCHCSGSGSLQCCGTGTLISVWFRVYICMFMNQHPFQVLRNMP
KEIMYGSVEMRGRHLIYCIWYDGHVGSIIIFNMSFGSALNYGLNNVIMCCYCE
NMAIRNCCARTRRLMLKAVGLIIVRETCDDPPICSSRTPRRQRLRALMERHPI
LFSEYESSRSHSTRLL"
2798. .3046
/gene="orf7"
2798. .3046
/gene="orf7"
/notes="This sequence comes from Fig. 1"
/codon_start=1
/protein_id="AAB37509.1"
/db_xref="GI:1699396"
/translation="ITVLDCEARLDDFSDGFIITDPLARQETWIIDPKSSRTN
ENISFKATRAERTYTVKAGGRLTRAGVKINKDT"
BASE COUNT 759 a 701 c 848 g 811 t
ORIGIN

Query Match 37.2%; Score 328.8; DB 16; Length 3119;
Best Local Similarity 63.1%; Pred. No. 9.3e-85;
Matches 526; Conservative 0; Mismatches 302; Indels 6; Gaps 1;
QY 46 cgatctcgtgtgtctcgcacacgacagaggtgacgctacacgtctcttcttctgagaca 105
```


	Query Match	36.1%	Score 319.2	DB 16	Length 35100
	Best Local Similarity	62.4%	Pred. No. 7e-82		
	Matches 520	Conservative 0	Mismatches 308	Indels 6	Gaps 1
QY	46	cqatctcggtgtctcgcgcgcactcgcacagttagggatcgtctacctccttttgagaca	105		
DB	33171	CAATCCAGTCGCTCCGCCATCATCCATTCGAGGCCCGCTTTACCAAGATCTGATGAT	33112		
QY	106	gaacccgcgtaccactactcgtggagatcattcgcgtgtcgcgcgaatgaacactttigaca	165		
DB	33111	GAACACAGGGTTCACTGACATGATCAACATCCCTGTTGCCGATTTGTATCATGTCAGAT	33052		
QY	166	atgcacaacgtgagttacgttcgcgcgaggtctctccctcgcagttggtgatttcgcgtatcag	225		
DB	33051	TATCATATGTAAGTTCTGTCGCGGATACCATGTCGCTGCTGGCTTTACCTGCTCCAA	32992		
QY	226	gaatggggttgtctccctgggatattgtctcaacgcgggagagcttgbatctcctgaggaag	285		
DB	32991	GAGTTCCAGTCCCTCGGATATGATCCTCACCACAGAGAAATAAAATTTTAAAGA	32932		

REMARK	(sites)	source
REFERENCE	4. (bases 4831 to 10470)	/organism="Human adenovirus type 12"
AUTHORS	Shu, L.M., Hong, J.S., Wei, Y.F. and Engler, J.A.	/db_xref="taxon:28282"
TITLE	Nucleotide sequence of the genes encoded in early region 2b of human adenovirus type 12	/function="inverted terminal repeat"
JOURNAL	Gene 46 (2-3), 187-195 (1986)	
MEDLINE	87106854	
REFERENCE	5. (bases 1 to 3957)	/citation=[5]
AUTHORS	van Ormondt, H. and Galibert, F.	/citation=[6]
TITLE	Nucleotide sequences of adenovirus DNAs	/citation=[7]
JOURNAL	Cur. Top. Microbiol. Immunol. 110, 73-142 (1984)	/citation=[8]
MEDLINE	85002829	/citation=[12]
REFERENCE	6. (bases 1 to 3957)	/replace="a"
AUTHORS	Shinagawa, M. and Padmanabhan, R.	
TITLE	Comparative sequence analysis of the inverted terminal repetitions from different adenoviruses	/citation=[5]
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 77 (7), 3831-3835 (1980)	/citation=[6]
MEDLINE	81054665	/citation=[7]
REFERENCE	7. (bases 1 to 3957)	/citation=[8]
AUTHORS	Sugisaki, H., Sugimoto, K., Takanami, M., Shiroki, K., Saito, I., Shimoto, H., Sawada, Y., Uemizu, Y., Desugi, S. and Fujinaga, K.	/replace="t"
TITLE	Structure and gene organization in the transformed Hind III-G fragment of Ad12	
JOURNAL	Cell 20 (3), 777-786 (1980)	
MEDLINE	81022638	
REFERENCE	8. (bases 1 to 3957)	
AUTHORS	Tolun, A., Alestrom, P. and Pettersson, U.	
TITLE	Sequence of inverted terminal repetitions from different adenoviruses: demonstration of conserved sequences and homology between SA7 termini and SV40 DNA	
JOURNAL	Cell 17 (3), 705-713 (1979)	
MEDLINE	80001962	
REFERENCE	9. (bases 1488 to 3861)	
AUTHORS	Kimura, T.	
TITLE	Structure and sequence analysis of the transforming region E1b of human adenovirus type 12	
JOURNAL	Sapporo Igaku Zasshi 52, 253-267 (1983)	
MEDLINE	10 (bases 20966 to 22966)	
REFERENCE	Kruijer, W., van Schaik, F.M., Speijer, J.G. and Sussenbach, J.S.	
AUTHORS	Structure and function of adenovirus DNA binding protein: comparison of the amino acid sequences of the Ad5 and Ad12 proteins derived from the nucleotide sequence of the corresponding genes	
TITLE	Virology 128 (1), 140-153 (1983)	
JOURNAL	83277521	
MEDLINE	11. (bases 24334 to 24703)	
REFERENCE	Engler, J.A. and van Bree, M.P.	
AUTHORS	The nucleotide sequence of the gene encoding protein IVa2 in human adenovirus type 7	
TITLE	Gene 19 (1), 71-80 (1982)	
JOURNAL	83054637	
MEDLINE	12. (bases 1 to 530)	
REFERENCE	Shibata, H., Zheng, J.H., Koikeda, S., Masamune, Y. and Nakanishi, Y.	
AUTHORS	Cis- and trans-acting factors for transcription of the adenovirus 12 E1a gene	
TITLE	Biochim. Biophys. Acta 1007 (2), 194-191 (1989)	
JOURNAL	89150250	
MEDLINE	13. (bases 1 to 34125)	
REFERENCE	Juttermann, R., Weyer, U. and Doerfler, W.	
AUTHORS	Defect of adenovirus type 12 replication in hamster cells: absence of transcription of viral virus-associated and L1 RNAs	
TITLE	J. Virol. 63 (8), 3535-3540 (1989)	
JOURNAL	89311650	
MEDLINE	14. (bases 1 to 34125)	
REFERENCE	Kimura, T., Sawada, Y., Shinagawa, M., Shimizu, Y., Shiroki, K., Shimoto, H., Sugisaki, H., Takanami, M., Uemizu, Y. and Fujinaga, K.	
AUTHORS	Nucleotide sequence of the transforming early region E1b of adenovirus type 12 DNA: structure and gene organization, and comparison with those of adenovirus type 5 DNA	
TITLE	Nucleic Acids Res. 9 (23), 6571-6589 (1981)	
JOURNAL	82105565	
MEDLINE	(sites)	
REMARK	Location/Qualifiers	
FEATURES		
1. 34125		
/organism="Human adenovirus type 12"		
/db_xref="taxon:28282"		
1. 161		
/function="inverted terminal repeat"		
6		
/citation=[5]		
/citation=[6]		
/citation=[7]		
/citation=[8]		
/citation=[12]		
/replace="a"		
8		
/citation=[5]		
/citation=[6]		
/citation=[7]		
/citation=[8]		
/citation=[12]		
/replace="t"		
9. 407		
/note="crossref Ad12 EPD30061 and EPD11197"		
/function="E1a promoter region (-499 to +100)"		
369		
/citation=[5]		
/citation=[6]		
/citation=[7]		
/citation=[8]		
/citation=[12]		
/replace="t"		
503. 1099		
/note="E1A; transcription activation: early protein; alternative splicing; crossref SWISS-PROT:E1A_ADE12, P03259"		
/codon_start=1		
/protein_id="CAA51877.1"		
/db_xref="GI:313362"		
/db_xref="SWISS-PROT:P03259"		
/translation="MRTETPLVLSYQAEADLLEHLVDNFNEVPSSDDLLYVPSLYEL YDLDVSGEDNNEQAVNEFFESLILASEGLFPEPVLSPVCEPIGECMPOLHP EDMDLICYEMGFPCSDSEDEQENAHVSASAAAANDREERFQLDHPQLPGHCK SCRHNSNTGNTDLMCSLCYLRYNNFIYKCAMGGGR"		
1039. 1628		
/note="crossref Ad12 EPD07152"		
/function="E1b promoter region (-499 to +100)"		
1542. 2033		
/note="crossref SWISS-PROT:E1BS_ADE12, P04492"		
/codon_start=1		
/product="E1b protein, small T-antigen"		
/protein_id="CAA51878.1"		
/db_xref="GI:313363"		
/db_xref="SWISS-PROT:P04492"		
/translation="MELETVLQSFQSRQLQVTSKNTSGFWYLFGLSTLSKVNRVK EDYREEFENLADCPGLLADLCYHLVFOEKVVRSLDFSSVGRVTASINFLATILDK WSEKSHLSWDYMLDYMSQLRAWLKRRCVCIYSLARPLTTPPLTQEEKEERPNAY VEK"		
1740		
/citation=[9]		
/replace="t"		
1742		
/citation=[9]		
/replace="g"		
1847. 3295		
/note="crossref SWISS-PROT:E1BL_ADE12, P04491"		
/codon_start=1		
/product="E1b protein, large T-antigen"		
/protein_id="CAA51879.1"		
/db_xref="GI:313364"		
/db_xref="SWISS-PROT:P04491"		
/translation="MERIPPELGLHAGLHVNAAVEGMAEEGLHLLAGAFDHAATA DVARGEGGAEPCGGVEVNEQVGEHLDGEGPCADDRCDDQKKEKSEKAAVLS RLVNLMSPLEIETVYWGELQDEQRGDHMLQYKYSFQKTHLWEPWEDMECAIKAF AKLARPCDSYRITKTIVTTSCTAIIINGAIVEVDTSRVAFRCMOGMGPGVVLGDG ITFINVRFAGDKFKIMFEANTCLVLHGVIFLNFNCSNVCESNKNVARGCIFYGCKWK		

```
LVGRPKSLVKKLFKCVLALIVEGDAHHRHNAASENACFVLLKGMALIKHNKMGV
VSDQTRFVTCADGNCHTAKTHVHVSRRHCWPCVDHNMFRCTIHLGRGMRPS
OCNFSHNMLEPEVFSRVCLNGVFDLSVELCKVIRYNDOTRHRRCRCEGSSHLEUR
PVLVNLSELRSDHLTSLCLRTDISSDEDDN"
2071..2073
/citation=[9]
/replacement="gcg"
2417..2418
/citation=[9]
/replacement="ga"
2479..2741
/citation=[9]
/replacement="atactgaat"
3374..3808
/note="crossref SWISS-PROT:HEX9_ADE12, P03284"
/product="hexon associated protein (Protein IX)"
/db_xref="GI:313365"
/db_xref="GI:313365"
/translation="MNSTQNNALFDGGVFSPLTSRLPYWAGVRQNVVGVSTVDGRP
VAPNSLTLYATIGPSPPLTAASAAASAASTARSMAADFSTYNHLASNAVTRAVR
EDILTMALKEITLTAOLELSQKVEELADATHTPAQPTQ"
complement(3844..5202)
/note="homologue to maturation protein Ad5, Ad2, Ad7;
Query Match 31.18; Score 275.4; DB 16; Length 34125;
Best Local Similarity 59.8; Pred. No. 3.7e-69;
Matches 481; Conservative 0; Mismatches 321; Indels 3; Gaps 1;
QY 68 ctccgtacagtaggagtgctaccctctttgagacagaaacccgcgtac---catac 124
DB 32277 CACCATACAAACAGTACCAGTACCGCCCTGTGAGGAAACAGTCAAAGCTACCTGAGTA 32218
QY 125 tggaggatccctcgtcgcgcgaatgtacacatttgacacatgcacacagtgagttacg 184
DB 32217 CTTCGGAATTTCTTTATGCGCTGAGTGAATAGTCTGACTTTACATAATGTAAGTGAGG 32158
QY 185 tgcaggtctccctcagctgaggtatcagctgattcaggaatgaggtgttccctggg 244
DB 32157 TAAGAGCATCTCTCATGTGTAGGTTTACAGTGTGACAGGAATGCCAATACCGTGGG 32098
QY 245 atatggttctaacgcgggagagctgttaacccgtgaggaagtgtatgcgctgtgctgt 304
DB 32097 ATATGATTTCAATGATGATGATGATGATTTATTTGAAAAAATACATGATGATGATG 32038
QY 305 gtgtgccaacattgatcatcacagagatgatcatcattgacaggttaccagctggctc 364
DB 32037 GTGTGCCACTATAATGTTGAAGTTACTCAATATTATACATGTCATGACGGTGGCTTA 31978
QY 365 tccactgtcattgttccagtcctcgggttccctcagctgagtgatagccgagcaggtttgg 424
DB 31977 TTTATTTGTCATGCCAGCTCCGGGTTCTACATGATGATGATGATGATGATGATGATG 31918
QY 425 ccagctggtttgaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 484
DB 31917 GAGCCTGGTGTAAATGGCGTATATGGCGCTTAATTAACAAAAGGTGTTTGGTATC 31858
QY 485 gggaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 544
DB 31857 GGGAGGTGTGTAAACATTTAATGCTTAAGAGGTGATGATGATGATGATGATGATGATG 31798
QY 545 ggggtgccaactatctacctgcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 604
DB 31797 GAGGTGGCCATTTAATTTACTTTAAATTTATGATGATGATGATGATGATGATGATGATG 31738
QY 605 cgcctatgagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 664
DB 31737 AAAAGTGTAGTTTGGATGGAGCGCTTTAATTTAATGGAATTTAATAAATGATGATG 31678
QY 665 tgtgtgcagttactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 724
DB 31677 TGTGTGTGTGATTTAAGAGCTTAAGTGTGATGATGATGATGATGATGATGATGATGATG 31618
```

```
QY 725 caaggcgccttatgctgcggcggtgcgaatcatcgtcagagaccactgcatgtgtgt 784
DB 31617 CCAGACTGCTAATGTTAAAGTTGTTCAAGTAATTCCTGAACACACTGTGCGCTCTAA 31558
QY 785 attctcgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 844
DB 31557 AACATAGTCGGCATGAACGTTATTCGTGACCAACTGCTAAAGGTTTAATATGATCATC 31498
QY 845 gccctatcctgacgacgattatga 869
DB 31497 GACCAATTTTATTTGAGATTATAA 31473
```

RESULT 8

```
ADERE4/c LOCUS 3498 bp DNA VRL 08-OCT-1999
DEFINITION Adenovirus type 12 DNA for early region E4.
ACCESSION X51800
VERSION X51800.1 GI:58543
KEYWORDS early region; early region E4.
SOURCE Human adenovirus type 12.
ORGANISM Human adenovirus type 12.
REFERENCE 1 (bases 1 to 3498)
AUTHORS Hogenkamp,T. and Esche,H.
TITLE Nucleotide sequence of the right 10% of adenovirus type 12 DNA
JOURNAL Nucleic Acids Res 18 (10), 3065-3066 (1990)
MEDLINE 90272430
REFERENCE 2 (bases 1 to 3498)
AUTHORS Hogenkamp,T.
TITLE Direct Submission
JOURNAL Submitted (06-FEB-1990) Hogenkamp T., Institute of Molecular
Biology (Cancer Research), University of Essex, Pufelandstr 55,
4300 Essen 1, F R G
COMMENT See <V00043> for overlapping sequence.
FEATURES
source
1..3498
/organism="Human adenovirus type 12"
/strain="Huie"
/db_xref="taxon:28282"
complement(816..1691)
/codon_start=1
/product="early E4 34 KD protein"
/protein_id="CAB57854.1"
/db_xref="GI:6018382"
/translation="MQRDRRYRLAPYKYLPPCEQSKATLSTSENSLWPECNSL
TLNHEVRGIPSCVGTFLQEMPFPDWMILTDEMFILKYNYSVMCCATINVEVTO
LHGHEDRLHCHQCPGSLQCSAGMLGRWFKMAYGALINKKRCFYREVNNLMP
KEYMYGTFVRGRHLIYFKIMYDGHAWLALEKVSFGSAFNYGILNNMLVLCDDYCK
DLSEIRMCWRPRRLHLLKVVQVIAENTVRLKHSRHERYRQQLKGLIMHRAILF
GYNORENPWADGH"
complement(1624..1986)
/codon_start=1
/product="early E4 13 KD protein"
/protein_id="CAB57853.1"
/db_xref="GI:6018361"
/translation="MPLPCIPPPVSRDTRACIAWLGLAHASCVDTLRFIKHDLKIT
PRAEVLASLREWLFPALTELRQRCQKRGATSGRTWFCFFKEDARKSVVYDAAR
QTVSLQIGLIQQVPTAL"
complement(1993..2343)
/codon_start=1
/product="early E4 11 KD protein"
/protein_id="CAB57852.1"
/db_xref="GI:6018360"
/translation="MKYCLRMVAGALTELFNINGLNQLONQCVQIIQWKNENYIGMY
QSGSLMIEFHDFNALFLFTEIRAVALLAEVVEHLENRLQDLAVIFHOHSGGDRCH
LRDLRIQLIADLRD"
```

BASE COUNT
ORIGIN

1131 a 864 c 620 g 883 t

	388	ggttcctcycagtgatagccgcggcgcaaggatttggccagctggttaggatggtggtg	447
y			
b	32385	GGGCTCTTCAATGTAAAGCGGGGGTGATGTACGAATAATGGTTTAAAATGATGATT	32326
y			
y	448	ggtgcgcacatgtttaatcacagagtttatatggttacccggaggggtgaatacacacatg	507
b	32325	TATGGTTCACGTATATAACTTAAGATTCCCTTGGTATCGAGAAATGTAAATTTGGCATG	32266
y			
y	508	ccaaaaggaggttaagttttatgccagcggttttatgaggggtgcaccattaactcacctg	567
b	32265	CCCAAGGAAGTGTATTTATGGGAGCGGTGTTCTCAGAGGTAGGCATTTATATATATT	32206
y			
y	.568	cacttgtatgatggcagcggtggttctgtgtcccccccatgagcttggatcacgc	627
b	32205	CGCTTCTGTATGATGGGCACGCGCTATAGTGTGGAGCGCATGTCGTTGGGTGGAGT	32146
y			
y	628	gccttgacctggtggatatttgaacaatatgtggtgtgtgtgcagttactgtctgat	687
b	32145	TTTTTTTAGTTATGGAATTATGAATATATGTTGTGTGTGGCTGTACTATTGTAAAC	32086
y			
y	688	ttaagtgcagatcaggtgcgtctgtgcccgaggacaaagcgcccttatgctgcgggg	747
b	32085	TTGGATGAGCTTCTCTGCGGTCTGCGCTCGTCGGACGCGCGCTGCTGGCCAAGCGC	32026
y			
y	748	gtcgaatcacctgctgagagaccactgccattgttgtattcctgcagggacggagcggcgg	807
b	32025	GTGAAGTGTGGGAAGCTTACCCTTAGTCATCATGAGAGCTTGTGGAAACCCCGG	31966
y			
y	808	cggcagcagtttatccgcgcgtctgcgcagcccaccccttctcctgatgcacgattat	867
b	31965	AGACAGGGATTACTCGGGCTTGATGGAGCGGTACTCCCCGTTTACGTTGGCAGACTAC	31906
y			
y	868	gac	870
b	31905	AAC	31903
<hr/>			
	RESULT	10	
	AB026117/c		
	LOCUS	AB026117	34094 bp DNA " VRL " 22-APR-1999
	DEFINITION	Porcine adenovirus 3 DNA, complete genome.	
	ACCESSION	AB026117	
	VERSION	AB026117.1	GI:4630864
	KEYWORDS	Porcine adenovirus serotype 3, complete genome	
	SOURCE	Larocque,D.	
	REFERENCE	Porcine adenovirus 3 DNA.	
	AUTHORS	Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.	
	TITLE	1 (bases 1 to 34094)	
	JOURNAL	Unpublished (1999)	
	REFERENCE	Larocque,D., Malenfant,F., Massie,B. and Dea,S.	
	AUTHORS	Direct Submission	
	TITLE	Submitted (14-APR-1999) to the DDBJ/EMBL/GenBank databases. Daniel	
	JOURNAL	des Prairies, Laval, Quebec H7N 4Z3, Canada	
		(E-mail:daniel.larocque@videotron.ca, Tel:(450) 687 5010)	
	FEATURES	Location/Qualifiers	
	source	1..34094 /organism="Porcine adenovirus 3"	
		/db_xref="taxon:35265"	
		/note="porcine adenovirus 3 strain IAF is a high titer clone, derived from strain 6618 and adapted for cell line PKA (pig kidney, porcine circovirus free)."	
	repeat_region	1..144	
	TATA_signal	449..452	
		/note="possible promoter region of ELA genes"	
	CDS	533..1222	
		/note="Description: ELA; homologous to ELA 289R of human adenovirus 2 and 5; putative early genes transactivator, induction of host cells DNA synthesis and activate p53 downstream events cascade"	

056..22912
 Note="major virion component"
 codon_start=1
 product="hexon"
 protein_id="BAA76968.1"
 xref="GI:4630875"
 translation="MATPSMPQSYMHISGDAEYISPLGVPSQATEFYFLNNK
 VTPVAPDVTTSRQLRPVYDKEDQYIKYKFLQAGDNRVLDMASTFFD
 LDKGSEKPYSGTAYNMAPKSNCCIDLPKGETBAGKYNVIAQAFVGPID
 IDLPYEQPLGSPSSANIDPSATSGAGRVLKQTPPRQY
 ASTNTHGGQTKDDKVTYFTFNNTATEAALNGLNPTVLSKEDVLDKAPT

```
RESULT 11
PAD237815/c 34094 bp DNA VRL 04-MAY-1999
LOCUS Porcine adenovirus serotype 3 complete genome, strain IAF.
DEFINITION
ACCESSION AJ237815
VERSION AJ237815.1 GI:4678602
KEYWORDS 100 kDa protein; 202R; 214R; 229R; 474R; 52 kDa protein; COOH
pix; proteinase; PTP; pVII; pVIII.
SOURCE
ORGANISM Porcine adenovirus 3
Virus; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
REFERENCE 1 (bases 1 to 34094)
AUTHORS Larocque,D., Malenfant,F., Massie,B. and Dea,S.
TITLE porcine adenovirus serotype 3, complete genome
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 34094)
AUTHORS Larocque,D.
TITLE Direct Submission
JOURNAL Submitted (05-FEB-1999) Larocque D., INRS-Institut Armand-Frappier,
331 boul. des Prairies, Laval, Quebec H7N 4Z3, Canada
FEATURES
source
1. 34094
/organism="Porcine adenovirus 3"
/strain="IAF"
/db_xref="taxon:35265"
/notes="high titer clone derived from strain 6618 and
adapted for cell line PKA (pig kidney, porcine circovirus
free)"
repeat_region
1. 144
/rpt_type="INVERTED"
/rpt_type="TERMINAL"
449..452
TATA_signal
/gene="E1A"
/notes="possible promoter region of E1A genes"
449..1291
/gene="E1A"
533..1222
/gene="E1A"
function="putative early genes transactivator, induction
of host cells DNA synthesis and activation of p53
dependent apoptosis cascade"
codon_start=1
/product="E1A 229R"
/protein_id="CAB41021.1"
/db_xref="GI:4678604"
/translation="MANRLHLDWDGNPEVVPVLEWDPVLDLDPSPGDEGFCPCWESL
VDGLPDELSDVDEVI VTEGESDSSGSAAGSGSQGYFEMDPPEGDSNEEDI
SAVAEVLSELADVFEDLPAPPSPVLDPCPEVGVNCRSCDIHRFHSKDPNLKSLC
YMRHAFVYGGCFWTFVGLCGKKGKSKSHVLFPIFCLFRSCFSSTSQVGPFTW
RPAQTEEVLL"
join(533..1043,1141..1274)
CDS
/codon_start=1
/gene="E1A"
/product="E1A 214R"
/protein_id="CAB41020.1"
/db_xref="GI:4678603"
/translation="MANRLHLDWDGNPEVVPVLEWDPVLDLDPSPGDEGFCPCWESL
VDGLPDELSDVDEVI VTEGESDSSGSAAGSGSQGYFEMDPPEGDSNEEDI
SAVAEVLSELADVFEDLPAPPSPVLDPCPEVGVNCRSCDIHRFHSKDPNLKSLC
YMRHAFVYGGCFWTFVGLCGKKGKSKSHVLFPIFCLFRSCFSSTSQVGPFTW
RPAQTEEVLL"
polyA_signal
1286..1291
/gene="E1A"
1461..2069
/function="putative inhibitor of apoptosis (Bcl-2
homologue)"
/codon_start=1
/product="E1B 202R"
/protein_id="CAB41022.1"
/db_xref="GI:4678605"
/translation="MDYQLLAKLTNNVLRKVIVQSGNCPWKKIFSDRFIKVVAEA
RRYQGLIEIFVEGERGFPEFLRGLGLEAVLKELDSTLGRTVASVALVCFE
KLOHSGWTDEGLSLVPLCLSLLEARMMAEQVRQGLCIIRMPASREMLLP6SGSG
SGSGAGMRDQVVPKPRPREQEEDEEDGMEASRRLEGPDIV"
1829..3253
/function="putative inhibitor of apoptosis by p53
dependent mechanism"
/codon_start=1
/product="E1B 474R"
/protein_id="CAB41023.1"
/db_xref="GI:4678606"
/translation="MPPAGGANDGGAAGAVHODAEERGAGDAVAOVVIRQWGRD.
AGPGAQAPAGAGRGGGGWDGSGRAQARAGSLDRRPGGAGGEGSGEAGGSSM
VSYQVLSLEYLSEPLEMHERYSEQIRPIMLQPGDDLGEMIAQHAQKAVLQGTVELR
RPTIRSMCYIIGNGAKIRGNTEYINIEPRNMCSTAGMWSVTITDVVDERLPA
RGLLIANTHFILHGCNFTLGSVITANAGGVVRCGYFCACIKALDRGLMLTYNE
NTEKCVYAVSAGRCRIKNSLSLSTFCFLHMSYTKIVGNSIMSPYTSDDPYDLV
CQSGCMPLSTVHIAPSSRLPYPEFRKNVLLRSTMFVGRGLSGSPRCSYSLSLV
VDEQSYRGLSVTCCEDOTCEMYKLLQCTEADMETDTSQYACLCGDNHPVQVRQMK
VTDALRAPKSLVSCNMGFEFSDDD"
3394..3993
/function="hexon-associated protein"
/codon_start=1
/product="pIX"
/protein_id="CAB41024.1"
/db_xref="GI:4678607"
/translation="MSSEGEIRTCFISARLPSPWAGVROGVAGTVNGVGVGAPAQSGV
LAYSREFVQQQQQPGTAATGVSFRAVPPSDLSAEVGMROALAELOQLREVVE
IQLRATASEAEEEEEEIIVDEEAPAGAGANTEEEEDEMLTMTVGDPEPAGVEA
QPPPTPTPESDPAPVATTTTPKRLSYGASKRSGPCAEEN"
4085..4090
/gene="E1B"
4085..4090
/gene="E1B"
polyA_signal
4085..4090
/gene="E1B"
CAAT_signal
6009..6015
complement(5962..5965)
TATA_signal
6009..6015
/gene="major late promoter (MLP)"
6009..6015
/gene="major late promoter (MLP)"
complement(8555..10396)
function="essential for DNA replication"
codon_start=1
/product="PTP (COOH region)"
/protein_id="CAB41025.1"
/db_xref="GI:4678608"
/translation="VELFRPVNINWRAHOWTRASVAVGIAMWSRFTYRHRMLMN
LSPROATEHPPLXLYPPHLLVGYOHLRLCNDYIFOTRAYSLRTEVTPHYQVY
NCPWATGTYINTGAYHRFLDLDSDLSQMOQAI LAERVADLALLRPLRGYGT
HDDADRVVPHLLQELQKIDRCQCEANGANDRLORAGRKDITLLRLQKLTAY
FHLVSRHLRPLTHALNLPCCCLDAFVERFAGSEORVEEAALQOQVPTRELT
RCIISALSLPCLSGTALRGAPFELRPGGGAATVETMRKRGVIERVYDLPLRRR
PRAPVPVAAAAAEEDEEMPELEGEAEAAVAGSTPPSPFEEVVRQVAENIAA
LEELTVSARQHEFFRAPEFYRVMQRDLGIDINELTLRRWVINFCEHIATTLNT
LHALLQLOPFNLVALNAQVVMRADQDQGVYSRVNMENGENALVIMNRITVL
AATVIRAGHGDLDDELEQFMODIARVNSGDVSEILRQVATHDADIDSVLSFRFY
TGPVFSFGHQIQINRRVIAHSLRDLPLPELNAPVPLAQLPA"
10629..11666
/codon_start=1
/product="52 kDa protein"
/protein_id="CAB41026.1"
/db_xref="GI:4678609"
/translation="MAPPAGALCAECATRSRPSLTPREEGEARLGAAGAAPER
HPVQLKREAMEAYVPKPAFRPGESEERDLRFAGRMQDRERVLPQDFEG
RVEEAGVSAARAHAASAAQIAEQTVREEVNFQKTFNNVTVTLVSRDEVMTGLHL
WDFVAFLOHPRSRALTAQIIAQHCDEGMVKKALLSLGASRWLVLDLYNLQTI
VQERSMSLSKVAALNYSVATLAKHYARKISTFYMRVAVVKKLVLLADLGMYNRKELE
RVYTSRRRELNDLELMFGLRRALAGEBDELEEDLEEELEERGGVRSPPGTAA
RENAVPAQDCR"
11718..13586
/codon_start=1
/product="I11a"
/protein_id="CAB41027.1"
/db_xref="GI:4678610"
/translation="MAASAGAAAAGVSAASLAPERATRMOALPSLDEPWEQALRRI
MAUTADUSRRFASQPLANRIGAILLEAVVPTNPTHEKVLTVVNNALLETSAIRPDEAG
```



```
CDS
19097..21916
/note="major structural protein"
/codon_start=1
/product="hexon"
/protein_id="AAC99441.1"
/db_xref="GI:4092659"
/translation="MATSMPQWSYMHISGQDASEYLSPLGVQFSQATETVFNLNK
FRNTPAHTDVTIERSQRLRFVVDKEDTQYTKYKRFQLAGNVRVLDMASTFFD
IRGLTDPGPFKPYSTAYNIMAKPSANNQYLDPKGTEAGKVNIAQASVFGPID
ETGTDKITTEDEDEETIDPLYEPQPLGPPSWSDNIPSATSGAGVLRKQTPROPCTY
GSASPTNIIGGQTKDDKTVLYFTNNPATEALEENGKLNKPVILISDVLKAPDT
HLVAVNTOEFAQYGLGQOAPNRYIGFRDNFTGLLIYNSNGOQMLAQOASQIN
AVYDLDNRHETSYQLDLSLHDSRIFSLWNOAIDSYDKDVRVLENNGVDEMFNC
FPGAETNMFTQLKRSNGGSRATTKENGDDGGNGAEHYLGIGNLNAEIMLTA
NWRSELYSNVLYLPDKYKFSPPNYPIDNTHSYDIKRLPLNNLIDTFVNICARW
SPMDNVNFIHNRNYGKYSQLLNGRYCKFHQVPOKFFALKSLIDLLPGAVTY
EWSFRDVMNLIAGTNDLRADGAKINTESNLYASFPMAHNTASILEAMLRNDTN
NOTFIDFLSANLXYIPANVTNLPISIPSRNWAARFGWFTRLKHNETPALGSPFDP
YFTYSGSIPVLDGTFYLGHTFERRISIQFSSVAMPNGDKLTPNEFEKRVVDGEGY
VAQNTKDKWFLYOMLAHNIGYQGHLPEDGRDRTYSLFRNFPEMCRQVPIYANHKD
EYLEVETNQFNSGFSVSAFTAGMRHGFYIPANWPLYIGEDAVQTVQRKFLCDRT
LWRIPSSNFMNGTITDLGQNLNLSAHALDMTFEVDAMDEPTLLVLEVEVDVCG
VHOPHGVIEAVLYLTPFSAGNATT"
21934..22548
/note="23k protein"
/codon_start=1
/product="proteinase"
/protein_id="AAC99442.1"
/db_xref="GI:4092660"
/translation="MGSTDELRAMARDLQLPFLGTPDKFGPGFLOESQRCQAIVNT
AARHTGGRWLAFAVAPASRIFYFPDPDFSDRELAQVYDFEYORLLRKSLSQSTPR
CITLYTQSGVGPASACGLFLCLFAAFRYPDSPMAYNPMDLVEGVNDKELFDA
DVQPIFRANOACYAFARHSAYFRAHRAHMEQTHLHKALDMQ"
24056..26572
/codon_start=1
/product="100k protein"
/protein_id="AAC99445.1"
/db_xref="GI:4092663"
/translation="MEDHSAASELSSGAAPLPPPPPPPTSPPPPSLQQQQPPT
ETDADERTCSSSSSSSECFVSPLETSSEDSADIVLSEPRDEEEDDESDPORY
MDADVLQRLHLLQTLRQVLEAAGAAEAAPKSVASLESLREALEAFSPATPPR
QENCTCAPDPLNYPYVPMLEALATYLLFHQKIPVSCRANPRADAHWRPSPGTP
LPDPTDVEYKIEGDEEPACANDKERDSVLVELKDNPRLAHVKQCIATVHF
APALAPKVMSTMQTLTVRRASPLDEGETPLEDLVVSDEQLARWMTSDPKVL
BERRTVACWMTVQLCHMTITSEWVRRLGECILHYMFGYVKLASKIANMELS
NLVSLGMHENRVLQVHHTLKHEARDYVREDTLYLYLYTWQIANGVWQCCLEDR
NLRALESLARQSLATGTPDETIAQDLAFLFPKLVETLQSLRSLPDAFASQSMHAF
RSFVLSRGILPVCNALPSDFVTYIRECPQPLWAHCYLLRLANFLMYHCDLAEDTS
GEGIFECYCRNLCAPHRCLATNALLNEVOAINTFELQRPKPDGTLPFPKLTPLGL
WTSGLRHFVSEYHSDRILFYEDVSRPPRPSACVITHSAILAOLHDIKKAREFL
LTKGHGYLDPHGTGEELNAAFTAHAAPEEAHPQOHOHQOQPSHRRHRHSYAD
RVRSGLHAYGATGSSRDPSVGGCSARGTHSRDAARRGSGQQRDQLRQRAQIPRG
TGGGGGTGHTDEAIOALLHQOQQQEHQPAQLRRPQRS"
27089..27760
/codon_start=1
/product="hexon associated protein precursor"
/protein_id="AAC99446.1"
/db_xref="GI:4092664"
/translation="MSKQIPTYWWSYQSGRAAGASVDSYSTRMNLWLSAGSMIGQV
NDIRTRNQILIRQALITETPRVQNPSPKASLLPQMTQPTHLLHLPNELEGRLT
DAGMLAGGALAPRDLXALTLEGRGIQNEDLPLASLTFRDGIQLGSGGSSFPNP
TDAYLILQNSSSLPRSGGSGSEQVREFVPTVIINPFGPPGTYPDQFIANYNLTDS
VAGYD"
27444..27447
TATA_signal
CDS
27753..28106
/codon_start=1
/product="putative 13.7 kDa E3 protein"
/protein_id="AAC99451.1"
/db_xref="GI:4092669"
/translation="MIDGPGQQRRELLDQHRQCPNRCCFAREGIHPEYFCITREH
FRAECIPDSLQSGHGLRFLSLTRYSDRRHRDGRDTILTYSYCGPASFKVRLCGHPAP
HPLDLLCELYNRS"
27976..28602
CDS
/note="putative 23 kDa E3 protein"
/codon_start=1
/product="23 kDa protein"
/protein_id="AAC99447.1"
/db_xref="GI:4092665"
/translation="METAPSLRRTAALLSKFAVSAAILLLTLFFSTFCVSCCTTARS
VSPTSRVRFLSDIEDSDSYSGSGDDEYELATNPNEIDILGSTVINNOIGP
TKALGYFYAAQMFVFFAIIIVILYRYRYVLATLIVQROMSSEAVLRKFTSATV
VTPPKQVTPCNGSCREEMVYITTSVFWPWSASSCSPPWSANPTG"
28307..28693
CDS
/codon_start=1
/product="putative 13.1 kDa E3 protein"
/protein_id="AAC99448.1"
/db_xref="GI:4092666"
/translation="MPPCSLSSPSSSSSSSTTAATCWPPSSCSARGCPRPSPCGK
PSRPPWLWLLPQNKSPPTATAPAAARRWCSTTPPSSCPGGPHPAHRHGGPPGLDSSGD
AOQEPAPAAATAPHLGTLRRGHLR"
28939..30285
CDS
/codon_start=1
/product="viral attachment protein"
/protein_id="AAC99449.1"
/db_xref="GI:4092667"
/translation="MGPKKQKRELPEDPVYDVPOLOINPPFVSGDGFNQSDVG
LSLHIAPLVFDNTRALTAFGGGLQSKQLVATEGLTTPDGLKLVLRKSPIT
LTAEGISLISGCLNSSETGLSLOVTAPOFGQNALTLPLAAGLQNTDGGVKRIGSG
LTTDSOAVTVQVNGLQNGEGQLTVPATAPLVSGAGISFNYSNDFVLDNDLSL
RKASVTPPLQSTEDTISLNSNDFSVNDGALTAPTKPTLTGASPTANVLTN
TTPNGFFCLTRTGGVGLVSGFALKSIDLTSMTKVNFIIDGAGRLQSDSTYKGRF
GRNSDVSLEPTAAGLSPALMPSTFIYPRNTSSGSLTSFVINTVYVVDIKYNTLS
TNGYSLEEFNQMSFSAFSTSYGTFCVPRTHRPHRGFSLRERHLQLLQQ"
30205..30929
repeat_region
30590..30940
CDS
/codon_start=1
/product="12.8 kDa protein"
/protein_id="AAC99450.1"
/db_xref="GI:4092668"
/translation="MTKKVNFIDGAGLQSDSTYKGRFGRNSDVSIEPTAAGLSA
WLMPSTFIYPRNTSSGSLTSFVINTVYVVDIKYNTLSNGYSLEFNQMSFSAF
STSYGTFCVYQPSA"
31183..31189
complement(31183..31189)
31305..31796
/note="ORF 8; similar to human adenovirus 34 kDa
polyptide"
/codon_start=1
/product="19.8 kDa polypeptide"
/protein_id="AAC99443.1"
/db_xref="GI:4092661"
/translation="MRSIMREIIRGCEYRIFWFRVAVNLPSVSRVMYVGSVMERGA
HLMYIKIMYDCLLRLRMFWHGEVYICDGYANYIVLCRRCHSELPVARRCAQRLR
HWLKLAEAIGAQRGRKQQTTEGWNWSRQDAEREERQALQECWLGSRVSQOI
KYF"
33398..33401
complement(33398..33401)
33768..33771
complement(33768..33771)
33951..34094
/note="inverted terminal repetition"
Query Match
Best Local Similarity 7.08; Score 62.2; DB 16; Length 34094;
Matches 317; Conservative 0; Mismatches 383; Indels 5; Gaps 2;
QY 163 acaatgcacaacgtgagtaacgtcaggtgttcctcgtcagtggtgatttcgatt 222
Db 32052 ACGTGTGATCCGGTGGGGGGCGCGCAACAGCGGCGCGGTATGTTTGTTCCTGT 31993
QY 223 caggaatgggtgttcctcgtgagatgttctaacggggagagagcttgaatcctgag 282
Db 31992 TTGGAGATTGTGGTCCCTCGCGCTACATCTTTCCCATCAAGAATGGATTTCTGCAG 31933
QY 283 aagtgtatgcacgtgtcctcgtgttggccaacatgatcatcatgcagcagcatgatc 342
Db 31932 TACAGCCTGTGTGGGGGGTGTGGAGCTGGAGCAGATATGTTTCAGCCCGGAGCGTGTGC 31873
```



```

AGATSTIYNPRKRYGAVVQVQATLSISAVLDNGOCVEIKYHSNLSASLTNLCNTNL
IDPLACLRPTTAHNPALIEEAAAPSLICYQVQVRAVEQAQVFLPFLKFLV
KGGVFLIKDSOMKCEFCFFKFTHTCALARRGFYFHHVAHSAWDLWISFTPI
GAPNTELFYVDEVYVTHWEGKGLVFLQGLDGDHSHVVCNPOLOEKATLIG
WNGEQDLYCYITPEKRAVGYKFRDFTLQHLIAASLSHSLVCONPOLOEKATLIG
LESPEELTPOOLKFLKNGPRFIEVAVGNITGFDEILLAAOVVSVKPMVPPEI
CRNMPRAGRLFPNDITYSLNPNFSPYPAKSYHEWQOVLSADLSQIKFMRVTFPS
LTHSLKNAKAYSLVSKGCPYQAVNEFVGLSGYQQDADGFPDLKTKWQOEEYSFN
KDMLEKKGAYDILQOTLDYCALDVQVTAQLVNLKLESYQIFKNSVNLPTFSFN
ORPTISSNAHIFKILYKAEPNTHSLTILMAWSENMEIEVRLSIRGRCYPTIG
VLOPVVYDIOGYSALHHPAGSPNPLTERALAKAYEQMKNHTKISYFDKDL
LPGFTTIDAPPBAEFLDVLPPCSKRGKGLWNEPLKEGATSIDVTLHNGWKV
TLPTDTRTTPPEKCLABEYVGLNLSAKEADKSNQTMRSIAKLNLNLYGSEAK
LONKTVSFSDIENLAKETAGAYVYKSSYIETDNLCAETMPEFVAVPVNSDVR
QLAPPSEEDPTOLPAEAPFMHNSMTSYHYKPIFIDEADDDCLHTELEKSTPLI
ANNRYQSASFLWTRAFVSEWQFLYENDAGTLENRLKSYGDTDSLEFTEG
YKMEERKRLKNGKGLVDFSPNPLTWLVECTQCEKGSAYSESIVLAPKLY
ALMDTTPCHVHGKGLRAKHATSTLSYDLVLCACYADMQGSDVFKTSMRLRT
LTSVQAHVQVFTVETTLTRKLRPKDKTLHALDMHRLIPYSRHPNPRNTETWMEL
complement(join(7941..9764,12575..12583))
/function="DNA replication"
/note="pnp"
/codon_start=1
/product="terminal protein precursor"
/protein_id="AAB38717.1"
/db_xref="GI:1732272"
/translation="MSLNALDCARLTGQTYTVEVFRPLNINFRVREYTRASTSVG
LAWMSKYIYQVRLMLNLSPREPATEGWFLWPPHLLVGYQYLVRTCDYVDFDR
SYRLKYTEIHLPLQKLNVTNANGSTINTGAYHRFIDENFAETLAOVQAVLAE
RVADLALIRPLAGYITNAGDQVPEVGLLDQHYKNSLQCNQAWGLDMRQQA
GNKDLVILTRKLTGTFNPLSPRNPTMLSCDCLNLDAFEKTPDPELSQFT
IQSLPSQVTSIIISALSRLPAPCTPLSGAFELPRENGRAVTEEMRRRGEVIER
FIDPLVRRRRRPPPPVSELSPEAEASPPPASPRASFEEDFVDFIVEAIRL
LOELTYSARNEQFFNAFVYEVQIORMGLNINELTIRRWMTVEFVAEIVATRL
LHNLRLYPPCSRWVLELAOVYMRARDEHQVYISKVWNEGENAFSQVMARYSGDL
AATVERAGLEGEEMEOPMADIAHENGSDVSELLROVAINDTVEVDSMELSRFV
9814..10980
/function="viral assembly"
/codon_start=1
/product="52.55kDa protein"
/protein_id="AAB38718.1"
/db_xref="GI:1732273"
/translation="MHPVLRMKPLPASAKLEAESEGLARLOGGAPELHPVOMKQ
EAEAVIPTSNVFRNEGEAEGLHAYESGRLLRNDHTSKRVIGRDPEKDPONGI
SAAEHLKSAADVAYEHTVKAENVFQFTFNNTVRLTIAREEVVIGLHMLDVEAPL
ENPVSAQLAQFLVQHCROGVNRESLNIADPEKRWLDVLLNLLQTLVQVGERGLA
VGEKVAINSVITLSKHYARKIFNSVFPIDKEAKINTFYMETVVKLLVLSDDLGM
RNERIAVSGARQRELNDRLMHLRLQALASNLPLEGEDSANISKEEWGVGAGG
GVASARYPHLLDYEEENPDGVSFQHERGAQPHENGHAESAYSRRQLGRFY"
10853..12556
/function="structural protein"
/note="pili"
/codon_start=1
/product="pili"
/protein_id="AAB38719.1"
/db_xref="GI:1732274"
/translation="MRKRTLMAPSHFNSMSALNPKMAAMQSOPTADDSSWAASIR
IMSLTAGDRHFKFSQPPANRLDALLAVVPSKDPHEKVLITVIALENGAIIRDEG
AGVYDALLHCKAYNSLNAQSNLERLAGVRAVQAVQVQRIATNGISGTALNGFLAR
PANYRGQENTGVFSALKLVSEVSTEVYQGHFYFLOVSRNGTQVNTNAPENL
KPLWGVQPMTEBLSITALLTPNLLLVLLVFFDTSISDSISVGLYLLTYREALG
RNHLDERTLEVTEVRASGSENINLQATLNLNRQKRPDKYSITPEERIVRF
VQAVSLRMQENLSTEDLDVTAANMESFYASNRDFINKMLDFFHRAAMADYFL
GAYNFWLPEFFGTFVDFPERDSYTWMDGLDLSLELTRODAMFEDLDFKDDQDT
ESKLSRVTTPASSRSSTVMPGGFMGMNNKNDLSREIDGLAKLAKWTKYKE
SEARVSLPAVVRPKOYRPSRPSISSDDSDGMSRDPPLKPEGSGNPFALRPLKRG
L"
12607..14040
/function="structural protein"
/note="viral component III"
/codon_start=1

```

CDS

```

Query Match
Best Local Similarity 5.68; Score 49.8; DB 16; Length 31323;
Matches 117; Conservative 0; Mismatches 112; Indels 0; Gaps 0
QY 356 cctgggtctccactgtcattgttccagtcctccggtccctccagtcagtcagcgcggc 415
Db 29026 CGTGGCGAGTACTGCTACTGCTCTAACAGCTGTCTTACATGCTAGCTAGCAAGC 28967
QY 416 aggttttggccagctgttagatggtggtgagtcgcccattgttaacagagttta 475
Db 28966 ATGNGGTTCAGAAAGTATGAGGACTTTTATAAGGGTGGGCTATGATAAAGATATA 28907
QY 476 tatggtaccggggagggtggaattacaacatgccaaagagtgatgttttccagcg 535
Db 28906 TGTGGTACCGAGAGTGTGAATTTCTAGCAGCCCTGATGAATTAACATATGCTGGGAGTA 28847
QY 536 tgtttatgaggggtgcacctaattaccctgcctgctgtggtatgatgg 584
Db 28846 TTATATTAGAACACTCACTACATATTTATTTAGACTGCTCTTTTATAG 28798

```

```

RESULT 15
CAUS5001/c
LOCUS
DEFINITION

```

```

CAUS5001 30288 bp DNA VRL
Canine adenovirus type 1 E1a protein, small T-antigen, large
T-antigen, protein IX, IVa2 protein, DNA polymerase, pTP, 52-55K
protein, pIIa, penton base protein, pVII protein, pV protein, Mu
peptide precursor, pVI protein, viral protein II, endoproteinase,
E2A-DBP, 100 kD protein, L4-22kD protein, L4-33 kD protein, pVII,
viral component IV, U-exon, E4 orf5, E4 orf4, E4 orf3, E4 orf2,
and E4 orf1 genes, complete cds.
U55001
U55001.1 GI:1477654

```

```

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

```

```

Canine adenovirus type 1.
Canine adenovirus type 1.
Viruses; GSDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
1 (bases 1 to 199)
Sira, S., Abouhaider, M.G., Liu, Y.C. and Campbell, J.B.
Multiple reiteration of a 40-bp nucleotide sequence in the inverted
terminal repeat of the genome of a canine adenovirus
Virology 159 (1), 76-83 (1987)
87265479
2 (bases 24577 to 27568)
Dragulev, B.P., Sira, S., Abouhaider, M.G. and Campbell, J.B.
Sequence analysis of putative E3 and fiber genomic regions of two
strains of canine adenovirus type 1
Virology 183 (1), 298-305 (1991)
91272490
3 (bases 1 to 30288)
Campbell, J.B. and Zhao, Y.
DNA sequence and genomic organization of canine adenovirus type 1

```

CDS

```

12607..14040
/function="structural protein"
/note="viral component III"
/codon_start=1

```

CDS

```

12607..14040
/function="structural protein"
/note="viral component III"
/codon_start=1

```

CDS

```

12607..14040
/function="structural protein"
/note="viral component III"
/codon_start=1

```


CDS

Search completed: June 23, 2000, 09:34:53
Job time: 12475 sec

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	885	100.0	3189	1	T31315	Adenovirus E4 and
C 2	885	100.0	3653	1	T59270	Plasmid pM17V4ORF
C 3	885	100.0	8299	1	T59273	Second generation
C 4	885	100.0	8299	1	T59273	Chimeric plasmid p
C 5	885	100.0	8710	1	V22144	Complete sequence
C 6	885	100.0	10610	1	V32370	Complete sequence
C 7	885	100.0	3026	1	T50559	Recombinant adenov
C 8	885	100.0	34303	1	T50559	Recombinant adenov
C 9	885	100.0	34382	1	X15627	Recombinant adenov
C 10	885	100.0	34427	1	X10371	Adenovirus vector
C 11	885	100.0	35000	1	T60557	Recombinant cis-ac
C 12	885	100.0	35935	1	V07258	Adenovirus 5 genom
C 13	885	100.0	36538	1	T60558	Recombinant trans-
C 14	883.4	98.8	1384	1	T31320	Glucocorticoid rec
C 15	875.4	98.9	36335	1	Q80003	A42/-ORF6/PGK-CFTR
C 16	370	41.8	36519	1	V22141	Chimpanzee adenov1
C 17	319.2	36.1	35099	1	V27112	Adenovirus 17. Rec
C 18	288.2	32.6	35524	1	V22140	Chimpanzee adenov1
C 19	49.8	5.6	6196	1	V13168	Complete DNA sequ
C 20	49.8	5.6	6243	1	V13841	Complete DNA sequ
C 21	49.8	5.6	6503	1	V13169	Complete DNA sequ
C 22	49.8	5.6	7379	1	V13176	Complete DNA sequ
C 23	49.8	5.6	8618	1	V18741	Complete DNA sequ
C 24	49.8	5.6	8792	1	V18745	Complete DNA sequ
C 25	37.8	4.3	34446	1	X24332	Complete DNA sequ
C 26	33.8	3.8	15894	1	V18267	Bovine adenovirus
C 27	33.4	3.8	502	1	Q65083	Measles virus Mont
C 28	33.4	3.8	509	1	Q65082	Human id-1H gene.
C 29	32.2	3.6	1391	1	V72242	Human DPl homologo
C 30	32.2	3.6	15894	1	V18264	Measles virus Edm
C 31	32.2	3.6	15894	1	V18268	Measles virus Rube
C 32	32.2	3.6	15894	1	V18269	Measles virus Mora
C 33	32.2	3.6	15894	1	V18270	Measles virus Zagr
C 34	32.2	3.6	15894	1	V18265	Measles virus 1977

1757 GCCTCCACGTGTCATGTTCCAGTCCCGGTTCCCTGCAGTGTATAGCCGCGGAGGTT 1698
421 ttggccagctgggttagatgagtggtggatggcgccatgttttaatacagaggttatgg 480
1697 TTGGCCAGCTGGTGTAGGATGGTGGATGGCGGCATGTTTATCAGAGGTTTATATGG 1638
481 taccggaggtggtgaattacaacatgccaaagagaggttaattgttccagcggtttt 540
1637 TACCGGAGGTGGTGAATTAACAATGCCAAAGAGAGGTAATGTTTATGTCAGCGTGT 1578
541 atgaggggtgcgccacttaatactacgtcgcttggatgagtgatggcgaggttctgtg 600
1577 ATGAGGGGTGCCACTTAATCTACCTGCGTGTGGTGTATGATGCCACGTTGGTCTG 1518
601 gtcccgccatgagcttggatagcagcgcttgcactgtgggatttgaacaattatgtg 660
1517 GTCCCGGCATGAGCTTTGGATATACAGCGCTTGACATGTGGGATTTGAACAATATGTG 1458
661 gtgctgtgctgagttactgtgctgatttaagtatgagatcaggggtgctgtgctgccc 720
1457 GTGCTGTGCTGAGTACTGTGCTGATTAAAGTGAATCAAGGTCGCTGTGTCGCCGG 1398
721 aggaacagcgcttactgagcggtgcgaatcctgctgagggagaccactgcccattg 780
1397 AGGACAAGCGCTTATGCTGCGGCGGTGGGATCATCGCTGAGGAGACCACCTGCCATG 1338
781 ttgtattctgcagagcgagcgcggtgagcagcttattatcgcgctgctgagcac 840
1337 TTGTATTCCTGAGGAGCGGCGGCGGAGCAGTATTATTCGCGCTGCTGTCAGCAC 1278
841 caccgcccattctgagcagcattatgacttactcctcccccattag 885
1277 CACCGCCCTATCTGATGACGATTATGACTCTACCCCATGTAG 1233

RESULT 6
V32375/c
ID V32375 standard; DNA; 10610 BP.
AC V32375;
DT 29-SEP-1998 (first entry)
DE Complete sequence of the pE4/Fiber plasmid.
KW Circular; adenovirus type 5; pE4/fiber plasmid; structural protein;
KW complementation; fiber protein; gene therapy; HIV; tumour; AD5;
KW early gene; Huntington's disease; Ray-Sachs disease;
KW sickle cell disease; E4 regulatory gene; ds.
OS Synthetic.
FH Key
FH misc_feature Location/Qualifiers
FT complement (21. .3149)
FT /tag= a
FT /note= "AD5 E4 regulatory gene"
FT sig_peptide 4051. .4366
FT /tag= b
FT /note= "AD5 leader sequence"
FT misc_feature 4372. .6124
FT /tag= c
FT /note= "AD5 fiber gene"
PN WO9813499-A2.
PD 02-APR-1998.
PF 24-SEP-1997; E05251.
PR 25-SEP-1996; US-719806.
PA (NOVS) NOVARTIS AG.
PA (SCRI) SCRIPPS RES INST.
PI Memerow GR, Von Seggern DJ;
DR WPI: 98-230709/20. - which lack DNA encoding for structural protein
PT Adenoviral vectors - which lack DNA encoding for structural protein
PT or fibre protein used particularly for gene therapy
PS Example 1; Pages 131-145; 170pp; English.
CC The present invention is that of a pE4/fiber plasmid used in the
CC method of the invention. The plasmid contains an adenovirus type 5 (AD5)
CC fiber gene controlled by a CMV promoter, an AD5 E4 gene and an adenovirus
CC type 2 (AD2) tripartite leader sequence upstream of the fiber gene. The
CC invention provides adenoviral vectors having deletions of all or part of

Huntington's disease; Tay-Sachs disease; sickle cell disease;
early gene; ds.
OS Synthetic.
FH Key
FH misc_feature Location/Qualifiers
FT complement (707. .3820)
FT /tag= a
FT /note= "E4 regulatory gene"
FT 3830. .6470
FT /tag= b
FT /note= "Hygromycin resistance gene"
PN WO9813499-A2.
PD 02-APR-1998.
PF 24-SEP-1997; E05251.
PR 25-SEP-1996; US-719806.
PA (NOVS) NOVARTIS AG.
PA (SCRI) SCRIPPS RES INST.
PI Memerow GR, Von Seggern DJ;
DR WPI: 98-230709/20.
PT Adenoviral vectors - which lack DNA encoding for structural protein
PT or fibre protein used particularly for gene therapy
PS Example 1; Pages 63-74; 170pp; English.
CC The present sequence is that of a pE4/Hygro plasmid used in the
CC method of the invention. The plasmid contains an adenovirus type 5
CC E4 regulatory gene and a hygromycin resistance gene. The invention
CC provides adenoviral vectors having deletions of all or part of
CC various gene sequences encoding adenoviral structural proteins.
CC Deletions in the structural proteins would allow a reduced risk of
CC wild-type virus contamination and would also allow packaging of foreign
CC DNA in such vectors for a variety of diagnostic and therapeutic
CC applications. The adenoviral vectors having deletions in the structural
CC gene regions are produced by cellular complementation of these
CC adenoviral genes. Therefore, the pE4/Hygro plasmid was used as a
CC complementation plasmid which was introduced into a host cell line
CC where parts of the E4 gene region would be stably inserted into the
CC host cell chromosomes. The resulting E4 deficient plasmid can be used
CC as a gene delivery vector. The vectors can be used for diagnosis or
CC gene therapy, e.g. for treating conditions characterised by
CC hyper-proliferative cells (e.g. tumours), genetic diseases (e.g.
CC Huntington's disease, Tay-Sachs disease, or sickle cell disease), or
CC infections (e.g. HIV infection). They can also be used for in vitro
CC production of biologically active proteins.
SQ Sequence 8710 BP; 2263 A; 2365 C; 2141 G; 1941 T;
Query Match 100.0%; Score 885; DB 1; Length 8710;
Best Local Similarity 100.0%; Pred. No. 1.1e-268;
Matches 885; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 atgactacgtccggttcatttggcagcactacgacacacacgactcgtgtgtct 60
DB 2117 ATGACTACGTCCGCGTTCATTTGGCATGACACTACGACACACACGATCGGTGTCT 2058
QY 61 cggcgactccgtacagtagggatgctctactcctctttgagacagaaacccgcgtacc 120
DB 2057 CGCGCACTCCGTACAGTAGGGATCGTCTACCTCCCTTTTGGAGACAAACCCCGCTACC 1998
QY 121 atactgagatcatcgctgctgcccgaagttaacactttgacaatgcacacgtgagt 180
DB 1997 ATACTGAGGATATCCGCTGCTGCGCGATGTAACACTTTGACATGACACACGTGAGT 1938
QY 181 tacgtcgaggtcttccctcagtggtggatttacgctgtattcaggaatgggtgttccc 240
DB 1937 TACGTGGAGGTCTTCCTCGAGTGTGGATTTACGCTGATTCAGGAATGGTGTGTCCC 1878
QY 241 tgggatattgttcaacgcygagagactgttaactcctgaggaagtgtatgcagctgtgc 300
DB 1877 TGGGATATGGTGTCTAACCGCGGAGAGCTTTGATCTTGAGGAAGTGTATGCACGTGTC 1818
QY 301 ctggtttgtgcaacattgatcatcagcagcagcatgatccatggtttacagagctctg 360
DB 1817 CTGTGTTGTGCCACATTGATATCATGACGAGCATGATGATCCATGTTACGAGTCTCG 1758
QY 361 gctctccactgtcattgttccagtcctcctcagtcctcctcagtcagtcagtcaggtt 420

Db	30168	ATGACTACGTCGGCGCTTCANTTGGCATGTCACACTACGACCAACAGCATCTCGGTGTCT	30109
Qy	61	CGCGCACTCGTACAGTAGGATCGCTACCTCTTTGAGACAGAAACCGCGCTACC	120
Db	30108	CGCGCACTCGTACAGTAGGATCGCTACCTCTTTGAGACAGAAACCGCGCTACC	30049
Qy	121	atactggagatcatccgctgctgcccgaattgaacactttgacaattgacaaagtga	180
Db	30048	ATACTGGAGGATCATCCGTCGTCGCCGAATGTACACTTTCACAATGCACAACGTGAGT	29989
Qy	181	tacgtgcaggtcttcctgcagctggtgggaattacacgtgattcaagaatgggtttcc	240
Db	29988	TACGTGCGAGGTCTTCCTGCAGTGTGGATTTACGCTGATTACAGAAATGGTGTCTCC	29929
Qy	241	tgggatattgttcaacgcggggagcttgtaactcctgaggaagtgtatgcacgtgtgc	300
Db	29928	TGGGATATGGTCTTACCGCGGGAGGAGCTTGTAACTCTGAGGAAGTGTATGCACGTGTC	29869
Qy	301	ctgtgtgtgccaacttataatgacgagcatgatcatcoattgattacagctccgg	360
Db	29868	CTGTGTTGTGCCAATGATATCATCAGCAGCATGATGATCATGGTTACGAGTCTCTGG	29809
Qy	361	gctctcaactgtaattgttccagtcctgggttccctcagtggtatagcgcggcgaggtt	420
Db	29808	GCTCTCCACTGTCAATGTGTCCAGTCCCGGTCTCCGAGTGTATGCCGCGGGCAGGTT	29749
Qy	421	tggcgacagctggtttagatggtgtggatggcgccatgtttaatcagagtttatagg	480
Db	29748	TTGGCCAGCTGTTTTAGGATGGTGTGGATGGCCCATGTTTAAATCAGAGGTTTATATGG	29689
Qy	481	taccgggaggtgtgaattacaacatgccaaaagagtaagtgttatgtccacgtgttt	540
Db	29688	TACCGGGAGGTGGTGAATTTACAACTGATGCCAATGCCAAGAGGTAAATTTATGTCACGCTGTT	29629
Qy	541	atgaggggtgcgcacttaattcacctgcgctgtgtgtatgtagccacgtgggtctctg	600
Db	29628	ATGAGGGTGCACACTTAATCTACCTGGGCTGTGGTATGATGACCAGTGGGTCTCTGTG	29568
Qy	601	gtccgcacatgagcttggatcacagcccttgcaactgggattttgaaacaattgtg	660
Db	29568	GTCCCCGCATGACCTTTGGATACAGCGCTTGCACTGTGGGATTTTGAAATATATGTG	29509
Qy	661	gtcgtgtgcagttactgtctgatttaagtggatgaagggtgcgtgtgtgtcccg	720
Db	29508	GTGCTGTGCTGCAGTTACTGTGCTGATTTAAGTGAGATCAGGGTCCGTGTGTGCCCGG	29449
Qy	721	agacaaggcgcttatgtcggggcggtgcgaatcatcgctgagagagaccactgcacgt	780
Db	29448	AGGAACAAGCGCCTTATGCTGCGGGCGGTGCGGATCATCGCTGAGGAGACCATGCCATG	29389
Qy	781	tgtattcttcgaggacggagcgggcgggcgagcagtttatcgcgcgtgtgtgagcac	840
Db	29388	TTGTATCTTCGAGGACGGAGCGCGGCGGACAGTATTATTTCGCGCGCTGTGTGAGCAC	29329
Qy	841	cacgcccctatcctgatgcagttatgactctacccccatgtag	885
Db	29328	CACCGCCCTATCTGTATGCAGATTTGATCTTACCCCATGTAG	29284

RESULT 8

V07261/C

ID V07261 standard; DNA; 34303 BP.

AC V07261:

DT 28-SEP-1998 (first entry)

DE Adenoviral vector plasmid pBHGL1.

KW Adenovirus 5; Ad5; vector; gene therapy; ds.

OS Mastadenovirus 5.

PH Key Location/Qualifiers

FT misc_feature 8772..9385

FT FT /note="a

FT FT /note="region deleted in plasmid pdeprp and

FT FT pBHGL1delPTP (Claim 45)"

FT FT misc_feature 8773..9586

FT		/tag= b	"region deleted in plasmid
FT		/note=	PAXBdelpoldeplpPVARNA+tl3 and
FT			pBHGlidelpolydeplpPVARNA+tl3 (Claim 53)"
FT			.11134
FT		c	"region deleted in plasmid pdePTP and
FT		/tag=	pBHGlidelpTP (Claim 51)"
FT		/note=	
FT	misc_feature	10705.	12513
FT		d	"region deleted in plasmid
FT		/tag=	PAXBdelpoldeplpPVARNA+tl3 and
FT		/note=	pBHGlidelpolydeplpPVARNA+tl3 (Claim 53)"
FT			
PN	WO9817783-A1.		
PD	30-Apr-1998		
PF	23-OCT-1997; UI9541.		
PR	23-OCT-1996; US-735609.		
PA	(UNMI) UNIV MICHIGAN.		
PI	Amalfitano A, Chamberlain JS, Hartigan-O'Connor DJ,		
PI	Hausner MA, Kumar-Singhr;		
PI	WPI: 98-261485/23.		
DR	New adenoviral recombinant plasmid(s) - comprise sequences provided		
PT	for expression of large foreign DNA fragments, used for, e.g. gene		
PT	therapy of genetic disease(s)		
PT	This nucleotide sequence comprises plasmid pBHGl1 that consists of		
CC	An E1-deleted adenovirus (Ad) genome. It contains a deletion of		
CC	Ad5 (see also V07258) from bp 188-1339 (0.5-3.7 m.u.). This		
CC	deletion removes the packaging signals as well as E1 sequences.		
CC	pBHGl1 also contains a large deletion within the E3 region (bp		
CC	27865-30995, 77.5-86.2 m.u.). Nucleotide 8773 of pBHGl1 is		
CC	equivalent to nucleotide 7269 of Ad5. The large E3 deletion		
CC	provides a large cloning capacity to Ad vectors based on pBHGl1.		
CC	Recombinant plasmids pdePol and pBHGlidelpol (deleted for E1 and		
CC	polymerase functions), and pAXBdelpoldeplpPVARNA+tl3 and		
CC	pBHGlidelpoldeplpPVARNA+tl3pBHGl1 (deleted for E1, polymerase as		
CC	preterminal protein functions), are specifically claimed. The		
CC	invention provides improved adenoviral vectors and packaging		
CC	cell lines. One type of improved vector comprises deletions		
CC	within the E2b region of the adenoviral genome (see also V07261).		
CC	These E2b-deleted virus are used in conjunction with novel cell		
CC	lines that constitutively express E2b gene products. The invention		
CC	also provides vectors deleted for all viral coding regions. These		
CC	'gutted' vectors permit the transfer of large genes (e.g. up to		
CC	35 kb) to cells, as demonstrated by the transfer of the dystrophin		
CC	gene to the muscle of mice. The E2b-deleted and gutted vectors		
CC	provide improved adenoviral vectors useful for a wide variety of		
CC	gene therapy applications.		
CC	Sequence 34303 BP;	9625 C;	9517 G;
SO			7213 T;

	Query Match	100.0%	Score 885	DB 1	Length 34303	
	Best Local Similarity	100.0%	Prod. No. 2.le-268			
	Matches 885	Conservative	0	Mismatches 0	Indels 0	Gaps 0
1	ataactacgtccggcgtttccatttggcagcactacacacacacacacgtctcgttgtct 60					
2						
3						
4						
5						
6						
7						
8						
9						
10						
11						
12						
13						
14						
15						
16						
17						
18						
19						
20						
21						
22						
23						
24						
25						
26						
27						
28						
29						
30						
31						
32						
33						
34						
35						
36						
37						
38						
39						
40						
41						
42						
43						
44						
45						
46						
47						
48						
49						
50						
51	cgcgcaactccgtacacgtaggagtcgtctacctccttttgacacagaacccgcgcctacc 120					
52						
53						
54						
55						
56						
57						
58						
59						
60						
61	cgcgcaactccgtacacgtaggagtcgtctacctccttttgacacagaacccgcgcctacc 120					
62						
63						
64						
65						
66						
67						
68						
69						
70						
71						
72						
73						
74						
75						
76						
77						
78						
79						
80						
81	atactggaggatcatccgtctgctgcgcgaatgtaacacatttgcacattgcacaacgtcgagt 180					
82						
83						
84						
85						
86						
87						
88						
89						
90						
91						
92						
93						
94						
95						

Db	32206	TGGGATATGGTTCTTAACGCGGGAGGAGCTTGTAACTCTGAGGAAGTGTATGACAGGTGTC	32147
QY	301	ctgtgtgtgtgcaacattgatcatcacagagcatgatcatcatcattggttacagatcctgg	360
Db	32146	CNGTGTGTGCACATTTGATATCATCAGCAGCATGATGATCATATGTTTACGAGTCTGTG	32087
QY	361	gtcttcactgtcattgttccagtcctcgggttccctcagtgatcagcggggcgaggtt	420
Db	32086	GCCTCCACTGCTCATTTGTCAGTCCCGGTTCCCTGCAGTGTATAGCCGGGGGAGGTT	32027
QY	421	ttggccagctgtgttaagatgtgtgtgtagtggcgccatgtttaatcagaggtttatatgg	480
Db	32026	TTGCCAGCTGGTTTAGGATGTGTGGATGCGGCATGTTTAAATCAGAGTTTATATGG	31967
QY	481	taccgggagtggtgtgaattacaacatgccaaagaggtaagtatttatgtccagcgtgttt	540
Db	31966	TACCGGAGGTGGTGAATTACAAATGCCAAAGAGGTAAATGTTTATGTCAGCGCTGTTT	31907
QY	541	atgagggttcgcaacttaactcactcgtcgttggtagtggccacgtgggtctcgtg	600
Db	31906	ATGAGGGGTGCGGCACATTAATACCTGCGCTTGTGGTATGATGCCAGTGGGTCTGTGTC	31847
QY	601	gtcccgcacatgagcttgttgatcacagcgccttgacatggaatttgaacaatatgtg	660
Db	31846	GTCCCGCCATGAGCTTGGATACAGCGCCTTGCACTGTGGGATTTTGAACAATATTGTG	31787
QY	661	gtactgtgtcagttactgtctgatttaagttagatcacgggtgcgctgctgtgccgg	720
Db	31786	GTGCTGTGTGCAGTTACTGTGCTGATTAAGTGGATCAGGNTGCGTCTGTGCCCGG	31727
QY	721	aggcaagcgccttatgtctggcgcggtgcgaatcatcgtcagagaccactgcatg	780
Db	31726	AGGACAAGCGCCTTATGCTGCGGGCGGTGGGAATCATGCTGTAGGAGACCACTGCCATG	31667
QY	781	ttgtattcctcgagcagcaggcggcggtggcagcagtttattccgcgctgctgtagaac	840
Db	31666	TTGTATTCTGTGAGACGAGCGCGGCGCAGAGTTTATTTCGCGCGTGTGTGAGCAC	31607
QY	841	caccgccttactcgtacgatttatgactctaccoccatgtag	885
Db	31606	CACCGCCCTATCCGTATGCAGCATTTAGTCTACCCCCCATGTAG	31562

RESULT 9

ID	Accession	Standard; DNA; 34382 BP.	Key	Location/Qualifiers
AD	X15627	standard; DNA; 34382 BP.	complement (3372..333)	
AC	X15627		/*tag- a	
DT	07-MAY-1999	(first entry)		
DE	Recombinant adenovirus Ad:Pac-beta-Gal.			
KW	Recombinant adenovirus Ad:Pac-beta-Gal;			
KW	Ela region; E3 region; gene therapy; nitric oxide synthetase; NOS;			
KW	cystic fibrosis chloride channel; Lbu receptor; erythropoietin;			
KW	atherosclerotic artery; ss.			
OS	Synthetic.			
OS	Mastadenovirus.			
HF	Key			
FT	CDS			
EN	US5880102-A.			
DD	09-MAR-1999.			
PF	17-JAN-1995; 374483.			
RR	17-JAN-1995; US-374483.			
PA	(UYDU-) UNIV DUKE.			
PI	Blazing MA, George SE;			
PR	WPI: 99-204005/17.			
RR	P-PSDB; W97243.			
TT	New replication deficient adenovirus bearing deletions of the Ela			
TT	and E3 regions - containing a single packaging signal sequence and			
TT	Ela enhancer sequence, the Ela deletion has unique cleavage sites			
TT	and is useful as a gene therapy vector.			
SS	Disclosure; Fig 49-100; 148pp; English.			
SS	The present sequence represents recombinant adenovirus			
CC	Ad:Pac-beta-Gal, which exemplifies the invention. The specification			
CC				

[illegible]

Db	34440	TGGGATATGGTTCTTAACGGGGAGGAGCTTGTTAATCTGAGGAAGTGTATGCACGTGTC	34381
QY	301	ctgtgttgccaaacatigatatcatgacagacataataatcoatgtttacagagtcctgg	360
Db	34380	CTGTGTGTGCCAACATTGATATCATGACGAGCATGATGATCATCGTTTACGAGTCTCG	34321
QY	361	gctctccaactgcattgttccactcccggttcctctcagtgatagccggcgggcaggtt	420
Db	34320	GCCTCTCACTGTCAATTGTCAGTCCGGTTCCTGCAGTGATAGCCGGCGGCGAGTT	34261
QY	421	ttggccacagctgttttagatggttgatggcgccatgttttaatacagaggtttatatgg	480
Db	34260	TTGGCCAGCTGGTTTAGATGGTGGATGGCGCCATGTTTAAATCAGAGTTTATATGG	34201
QY	481	taccgggagtggtgtaattacaacatgccaaaagaggttaattttatgtccagcgtattt	540
Db	34200	TACCGGGAGGTGGTGAATTACAACATGCCAAAGAGGTAAATGTTATGTCACAGCTGTTT	34141
QY	541	atgagggtgcgccacttaatactacgtcgcttggtatgatggccacgtgggttctgtg	600
Db	34140	ATGAGGGTGCGCCACTTAATCTACCTGCGTGTGTGATGATGGCCAGCTGGTCTCTGTG	34081
QY	601	gtccccgcacatgagcttttgatacagcgcttgcaactggggatttgaaacaattatgtg	660
Db	34080	GTCCCGCCCATGAGTTTGATACAGCGCTTGCACTGTGGGATTTTGAACAAATTATGTG	34021
QY	661	gtgctgtctgcagtactgtctgtatttaagtggagatcagggtgcgctgctgtgcccgg	720
Db	34020	GTGCTGTGCTGCAGTTACTGTCTGATTTAAAGTGAGATCAGGGTGCCTGCTGTGCCCG	33961
QY	721	aggacaaggcccttaigtgcggcggtgcgaatcatcgctgaggagacaactgcacatg	780
Db	33960	AGGACAGGCGCCTTATGCTGCGGGCGGTGCGAATCATCTGCTGAGGAGACCACTGGCCATG	33901
QY	781	ttgtattcttcgaggacggagcgcgcgcgagcagatttattcgcgcgtgctgcagcac	840
Db	33900	TTGTATTCTTCGAGGACGGAGCGGGCGGACAGAGTTTATTCGCGCGCTGCTGCAGCAC	33841
QY	841	caccgcctatcctgatgcacgattatgactacccccattgtag	885
Db	33840	CACGCGCTATCTGTATGCACGATTATGACTCTACCCCATGTAG	33796

RESULT 14

T31370
ID T31320 standard; cDNA; 1884 BP.
AC T31320;
AD 16-APR-1997 (first entry)
DE Glucocorticoid receptor/adenoviral ORF6 chimaeric fusion gene.
KW Human adenovirus; inverted tandem repeat; E4; E1; open reading frame;
KW expression; late gene; cell line; defective recombinant adenovirus;
KW complementation; gene therapy; cystic fibrosis; dystrophy; ds; PCR;
KW neurodegenerative disease; blood coagulation disorder; viral infection;
KW primer; amplification; polymerase chain reaction; fusion protein.
KW Homo sapiens/Adenovirus Ad5.
OS Homo sapiens/Adenovirus Ad5.
SH Key
TS Location/Qualifiers
CT 1. 1605
TT /*tag-a
TT /product= glucocorticoid receptor/adenovirus Ad5 ORF6
TT product fusion protein
TN W09622378-Al.
25-JUL-1996.
19-JAN-1996; F00088.
20-JAN-1995; FR-000747.
01-JUN-1995; FR-006532.
08-SEP-1995; FR-010541.
A (RHON) RHONE POULENC RORER SA.
A Dedieu J, Latta M, Orsini C, Perricaudet M, Vigne E;
I yah P, Dedieu JF;
I WPI; 96-354535/35.
P-PSDB: W03197.
T Cells for prodn. of recombinant adeno and adeno-associated virus -
T contain part. of the E4 region of adeno virus, used to produce new

Mon Jun 26 09:10:29 2000

```
QY 661 gtgctgtgtcgcagttactgtgatttaagtgaagtcagggcgctgtgtgcccgg 720
DB 1381 GFGCTGTGCTGCAGTTACTGTCTGATTAAGTGAATCAGGCTGCTGTGCTGCCGG 1440
QY 721 aggacaagcgcccttactgtcgccggtggaatcatcgtgagagaccactgcacatg 780
DB 1441 AGGACAAGCGCCCTTACTGTGCGGCGGTGCGAATCATCGCTGAGAGACACACGCCCATG 1500
QY 781 ttgtattctcgcagacgagcgcgcgcgagcagctgtatttattcgcgcgctgtgcagcac 840
DB 1501 TTGTATTCTCGAGACGCGGCGGCGGCGGCGAGCAGTTATTTCGCGCTGCTGCAGCAC 1560
QY 841 caccgcccctactcgtgcacgattatgactctaccctccatag 885
DB 1561 CACCGCCCTATCTGATGACGATTATGACTCTACCCCATCTAG 1605

RESULT 15
Q68003/c
ID Q68003 standard; DNA; 36335 BP.
AC Q68003;
DT 26-MAR-1996 (first entry)
DE Ad2/ORF6/PGK-CFTR nucleotide sequence.
KW Recombinant adenovirus; Ad2/CFTR-1; adenovirus 2 serotype; Ela; Elb;
KW viral replication; gene expression; gene therapy; cystic fibrosis;
KW cystic fibrosis transmembrane conductance regulator; CFTR;
KW promoter; E3; p19; MHC; class 1; viral latency; pulmonary airway; ds.
OS Synthetic.
FH Key
FT misc_feature
FT 12915..36335
FT /tag= a
FT /note= "Represents residues 10676-34096 of Ad2-E4/ORF6"
FT 35069..35973
FT /tag= b
FT /note= "Represents residues 33178-34082 of Ad2"
FT 12915..35054
FT /tag= c
FT /note= "Represents residues 1-32815 of Ad2"
FT 28478..28790
FT /tag= d
FT /number= 1
FT /note= "33K protein"
FT 28791..28992
FT /tag= e
FT /number= 1
FT 28993..29366
FT /tag= f
FT /number= 2
FT /note= "33K protein"
FT 13279..14526
FT /tag= g
FT /product= 52.55K protein
FT 14547..16304
FT /tag= h
FT /product= Iiia protein
FT /note= "peripentonal hexon-associated protein"
FT 16331..16336
FT /tag= i
FT /note= "Major late mRNA L1 poly A signal (putative)"
FT 16390..18105
FT /tag= j
FT /product= Penton protein
FT /note= "virion component III"
FT 18112..18708
FT /tag= k
FT /product= Pro-VII protein
FT /note= "precursor to major core protein"
FT 18778..19887
FT /tag= l
FT /product= pV protein
FT /note= "minor core protein"
FT 20180..20193
FT /tag= m
FT poly_a_signal
```

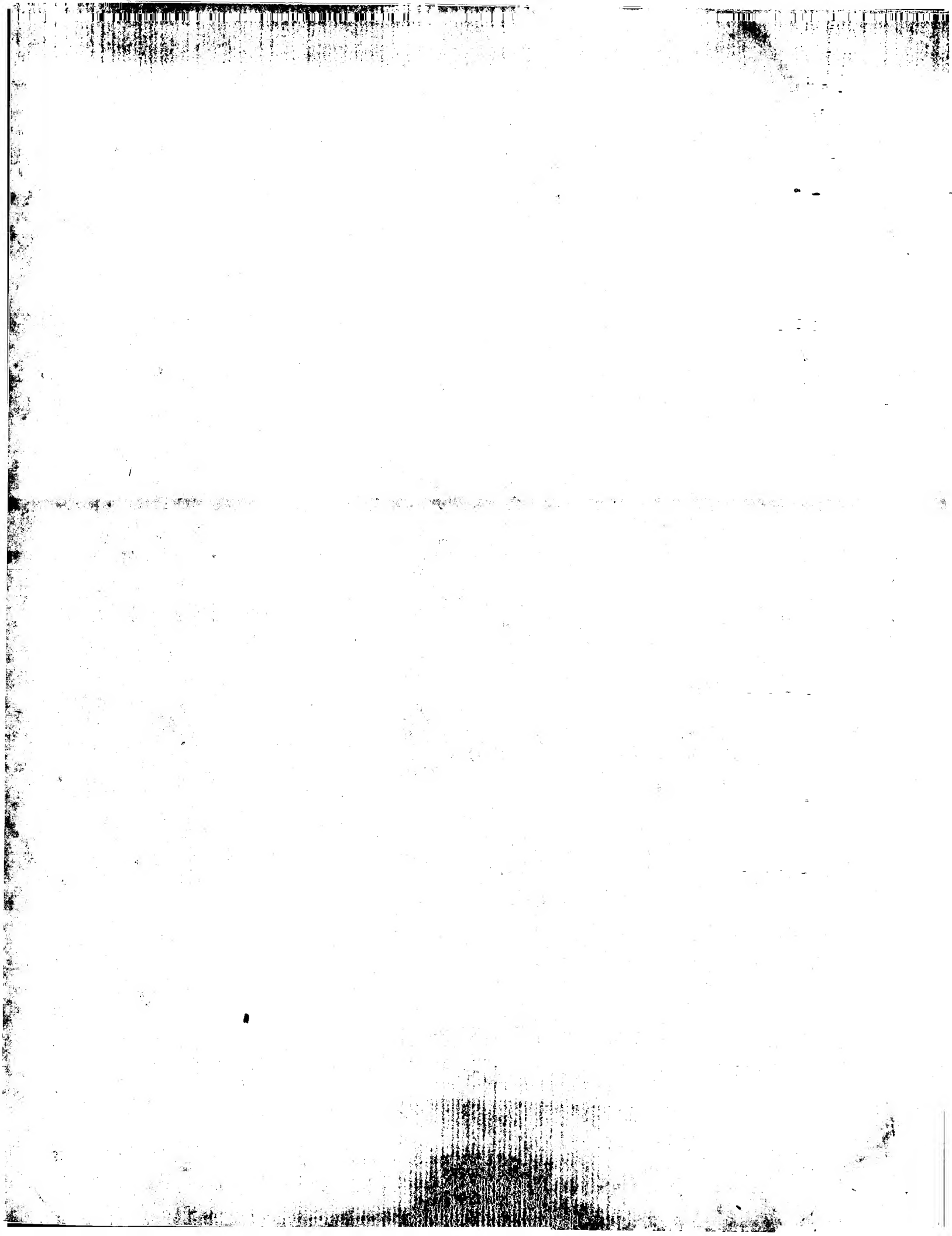
```
FT cds
FT 20240..20992
FT /tag= n
FT /product= pVI protein
FT /note= "Hexon associated precursor"
FT 21077..23983
FT /tag= o
FT /product= Hexon protein
FT /note= "Virion component II"
FT 24657..24662
FT /tag= p
FT /note= "Major late mRNA L3 poly-A signal (putative)"
FT complement (244729..26318)
FT /tag= q
FT /product= DBP protein
FT /note= "DNA binding or 72K protein"
FT 26347..28764
FT /tag= r
FT /product= 100K protein
FT /note= "Hexon assembled"
FT 29454..30137
FT /tag= s
FT /product= pVIII protein
FT /note= "Hexon-associated precursor"
FT 30444..30449
FT /tag= t
FT /note= "Major late mRNA L4 poly-A signal (putative)"
FT 31051..31530
FT /tag= u
FT /product= E3 19K protein
FT /note= "Glycosylated membrane protein"
FT 31707..32012
FT /tag= v
FT /product= E3 11.6K protein
FT 32008..32013
FT /tag= w
FT /note= "E3-1 mRNA poly-A signal (putative)"
FT 33081..33086
FT /tag= x
FT /note= "E3-2 mRNA poly-A signal (putative)"
FT 35013..35018
FT /tag= y
FT /note= "Major late mRNA L5 poly-A signal (putative)"
FT 1..12914
FT /tag= z
FT /note= "Represents residues 1-12914 of pAd2/PGK-CFTR"
FT 380..914
FT /tag= aa
FT /note= "pgk promoter"
FT 1011..5453
FT /tag= ab
FT /product= CFTR
FT W09412649-A.
FT 09-JUN-1994.
FT 02-DEC-1993; U11667.
FT 03-DEC-1993; US-985478.
FT 01-OCT-1993; US-130682.
FT 13-OCT-1993; US-136742.
FT (GENZ ) GENZYME CORP.
FT Armentano D, Couture LA, Gregory RJ, Smith AE;
FT WPI: 94-200277/24.
FT Adeno:virus-based gene therapy vectors - esp. useful for gene
FT therapy of cystic fibrosis.
FT Example 15; Page 84-95; 167pp; English.
FT This sequence represents the nucleotide sequence of the second generation
FT adenovirus vector named Ad2-ORF6/PGK-CFTR. This virus lacks EI and in its
FT place contains a modified transcription unit with the phosphoglycerate
FT kinase (PGK) promoter and a poly A addition site flanking the cystic
FT fibrosis transmembrane conductance regulator (CFTR) cDNA. The PGK
FT promoter is only if moderate strength but it is long lasting and is not
FT subject to shut off. The E4 region of the vector has also been modified
FT in that the whole sequence has been removed and replaced by ORF6, the
FT only E4 gene essential for growth of adenovirus (Ad) in tissue culture.
FT CC
```

CC The DNA construct comprises a full length copy of the Ad2 genome from
CC which the early region 1 genes (E1 genes) have been deleted and replaced
CC by an expression cassette encoding CFTR. The expression cassette
CC includes the promoter for PGK and a poly-A addition signal from the
CC bovine growth hormone gene. The Ad2-ORF6/PGK-CFTR construct differs
CC from that given in Q68002 (Ad2/CFTR-1). In that the latter utilises the
CC endogenous E1a promoter, has no poly-A addition signal directly
CC downstream of CFTR and retains an intact E4 region. This adenovirus may
CC be administered to the pulmonary airways in the gene therapy of cystic
CC fibrosis.
SQ Sequence 36335 BP; 8597 A; 10000 C; 9786 G; 7952 T;

Query Match 98.9%; Score 875.4; DB 1; Length 36335;
Best Local Similarity 99.3%; Pred. No. 2.3e-265;
Matches 879; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 1 atgactacgtccggcgttcatttgcatgacactacgacacacagatctcgggtgtct 60
Db 35968 ATGACTACGTCCGGCGGCTTCATTGGCATGACACTACGACCAACACGATCTCGGTGTCT 35909
QY 61 cggcgcactccgtacagtagggatcgtctacctctcttttgagacagaaacccgcctacc 120
Db 35908 CGCGGCACCTCCGTACAGTAGGATCCCTACCTCTTTGACAGAGACCCGCGCTACC 35849
QY 121 atactggagatccctcgtcgtgcccgaatgaacactttgacaacacgtgagt 180
Db 35848 ATACTGGAGATCATCCGCTGCTGCCGAATGAACACTTTGACAAATGAGTGTGAGT 35789
QY 181 tacgtgcgaggtctccctgcagtggtggattacgctgattcaggaatgggtgtctcc 240
Db 35788 TACGTGCGAGGTCTCCCTGCAGTGTGGATTACGCTGATTCAGGAATGGGTGTCTCC 35729
QY 241 tgggatattgtctaacgcgggagagcttgtaacctcctgaggaagtgatcacgtgtgc 300
Db 35728 TGGGATATGGTTCTACGCGGGAGGAGCTGTGTAATCTCTGAGGAAGTGTATGCAGTGTGC 35669
QY 301 ctgtgttggcacaatgatcatgacagcatgatccatggttaacagatccttg 360
Db 35668 CNGTGTGTGCCAATGTGATATCATGACGAGCATGATGATCCATGTTACGAGTCTCTGG 35609
QY 361 gctctccactgcatgttgcagtcctccggttccctgcagtgatagccggcgaggtt 420
Db 35608 GCCTCCACTGTCAATGTTCCAGTCCCGGTTCCCTGCAGTGCATAGCGGGGCGAGTT 35549
QY 421 ttggccagctggttaggatggttggtggatggcgcctatgtttaaacsagaggtttatgg 480
Db 35548 TTGGCCAGCTGTTAGGATGGTGGATGGGATGGGCCCATGTTTAAATCAGAGGTTTATATGG 35489
QY 481 taccgggaggttggtgaattacaacatgcacaaagagtgatgtttatgtccagcgtgtt 540
Db 35488 TACCGGAGGTGGTGAATTACAAATGCCAAAGAGGTAATGTTATGTCCAGCGGTGTTT 35429
QY 541 atgaggggtgcacacttaactacactgcgttggatgagatgagcgcgtgggtttctgtg 600
Db 35428 ATGAGGGGTGCCCACTTAATCTACTCGCTTGTGGTATGATGCCACGTGGGTTCTGTG 35369
QY 601 gtcccccgcactgagcttggatcacagcgcctgcactgtgggattttgaacaattatgtg 660
Db 35368 GTCCCCGCATGAGCTTGGATACAGCGCCTTGACCTGTGGGATTTGAACAATATTGTG 35309
QY 661 gtgctgtcgtcagttactgtctgtatttaagttagatcaggtgctgctgctgcccgg 720
Db 35308 GTGCTGTGCTGCAGTACTGTCTGATTTAAGTGAGATCAGGGTGGCTGCTGTGCCCG 35249
QY 721 aggacaaggccttatgtcggggcggtgcgaatcatcgtcgtgagagaccactgcatg 780
Db 35248 AGACAAAGGCGTCTCAGTCGCGGGCGGTGCGAATCATGCTGAGGAGACCATGCGCATG 35189
QY 781 ttgtattcctgcaggagcgggcgggcgagcagctttatcgcgcgtcgtcagac 840
Db 35188 TTGTATTCTTCAGGACGAGCGGGCGGCGGACGAGCTTTATTCGCGCGCTGCTGCAGCAC 35129

QY 841 caccgcctctctgtagtgcagcagattatgactctacccccatgtag 885
Db 35128 CACCGCCCTATCTCTGATGACGATTATGACTTACCCCATGTAG 35084

Search completed: June 23, 2000, 10:05:42
Job time: 3610 sec




```

; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States Of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/735,609
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: UM-02484
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-9410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 34303 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "DNA"
US-08-735-609-4

Query Match 100.0%; Score 885; DB 4; Length 34303;
Best Local Similarity 100.0%; Pred. No. 7.1e-282;
Matches 885; Conservative 0; Mismatches 0; Indels 0; Gaps

QY 1 atgactacgtccggcggttcaccattggcatcacacccaacacacatcfcgggttgtct 60
Db 32446 ATGACTACGTCCGGCGGTTCCATTGGCATGACACTACGACCAACACCATCTCGGTGTCT 32988
QY 61 cggcgcaactcgtacagtaggacgtctactccttttgacagacagaaaccgcgtacc 120
Db 32386 CGCGGCACCTCGTACAGTAGGATCGTCTACTCCTTTGAGACAGAAACCCGCGTACC 32322
QY 121 atactggaggatcgcgtcgtgcgccgaatgaacatttgacaaatgcacacgtgagt 180
Db 32326 ATACTGGAGGATCATCGGTGCTGCCGAATGTAACACTTGTACAATGCACACGTGAGT 32266
QY 181 tacgtcggaggtctccctgcagttgggattacgtgattcaggaaatgggttgtctccc 240
Db 32266 TACGTGCGAGGTCCTCCCTGCGAGTGGGAATTTACGTGATTGAGGAATGGGTGTGTC 32207
QY 241 tgggatatggttctaacycgggaggagctgttaactcctgaggaagtgtatgcacgtgac 300
Db 32206 TGGGATATGGTTCTTAACCGGGAGGAGCTTGTAATCTGAGGAAGTGTATGACGTGTGC 32147
QY 301 ctgtgtgtgcacaattgatatacatcacagagatgatccatggttacgagttcgtgg 360
Db 32146 CTGTGTGTGCCCACAACTTGATATCATGACGAGCATGATGATCCATGTTACGAGTCTGG 32087
QY 361 gctctccacgtgcattgttcaggtcccggttcctcagtgatagccggcgggcaggtt 420
Db 32086 GCTCTCCACTGTCATGTTCAGTCCCGGTTCCCTGCGAGTGTATACCGCGGGCAGGTT 32027
QY 421 ttggccagctggtttagatggttggtgattggcgccattgttaatcacagagtttatgg 480
Db 32026 TTGGCCAGCTGGTTTAGATGGTGGTGGATGGGCCCATGTTTAATCAGAGGTTTATAGG 31967
QY 481 taccggggagggtgggaattacaacatgccaaaagagtaattgtttatgtccaagcgtgtt 540
Db 31966 TACCGGGAGGTTGGTCAATTACACATGCCAAAAGAGGTAAATGTTTATGTTCAGCGGT 31907
QY 541 atgaggggtcgccaaactaaatactacgtgcgttggtagtatggccacagtgagttctatg 600

```

```
Query Match      100.0%; Score 885; DB 4; Length 34303;
Best Local Similarity 100.0%; Pred. No. 7.1e-282;
Matches 885; Conservative 0; Mismatches 0; Indels 0;
```

Db 31906 ATGAGGCGTCCCACTTAATCTACCTCGCTTGTGATATGATGCGCCACGCGGGTCTGTG 31847
Qy 601 gtccccccatgagcttgatagacagcgcccttgcaactgtgggattttgaaacaattattgtg 660
Db 31846 GTCCCCGCAATGACCTTTGGATACAGCGCCTTGACACTGTGGGATTTTGAACAATATTGTG 31787
Qy 661 gtctgtgtcagcttactgtctgatttaagtgtgagatcaggggtgctgtgtgccccgg 720
Db 31786 GTGTGTGCTGCTGCTACTGTGCTGATTTAAGTGTGAGATCAGGGTGCCTGTGTGCCCGG 31727
Qy 721 aggcacagcgcccttattctgctggcggtgcaaatcatcgtgagagacactgcccattg 780
Db 31726 AGGACAGCGCCCTTATGCTGCGGCGGTGGGAATCATCTGATGAGGACACTTGCCCATG 31667
Qy 781 ttgtattcttcagagcagcgagcgcggtggcgagcagctttattctcgcgctgtcagcac 840
Db 31666 TTGTATTCTTCAGGACGAGCGGCGGCGGAGCAGATTTATTTCGCGCGCTGCTGCAGCAC 31607
Qy 841 caccgcccctactctgacagattactgacttaccctcccatgtag 885
Db 31606 CACCGCCCTATCCTGATGACGATTTATGACTCTACCCCATGTAG 31562

RESULT 3
US-08-374-483-6/c
; Sequence 6, Application US/08374483
; Patent No. 5880102
; GENERAL INFORMATION:
; APPLICANT: GEORGE, SAMUEL E.
; APPLICANT: BLAZING, MICHAEL A.
; TITLE OF INVENTION: ADENOVIRAL VECTOR SYSTEM
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESS: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD, 8TH FLOOR
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/374,483
; FILING DATE: 17-JAN-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: WILSON, MARY J.
; REGISTRATION NUMBER: 32,955
; REFERENCE/DOCKET NUMBER: 1579-83
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; TELEX: 200797 NIXN UR
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 34382 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-374-483-6
Query Match 100.0%; Score 885; DB 3; Length 34382;
Best Local Similarity 100.0%; Pred. No. 7.1e-282;
Matches 885; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 atgactacgtcggcggttccatttgagatgacactacacacagatctcgtttctct 60

Db 32278 ATGACTACGTCCGGCTTCCATTTGGCATGACACTACGACCAACAGATCTCGGTGTGCT 32219
Qy 61 cggcgcaactcgtacagtaggactgtactctctcttttgagacagaaacccggcgctacc 120
Db 32218 CGGCGACATCCCTGACAGTAGGATCTGTACTCTCTTTTGGAGACAGAAACCCGCGCTACC 32159
Qy 121 atactggagagatccctcgtcgtcccaaatgtaacactttgacaattgacaacatgagt 180
Db 32158 ATACTGGAGGATCATCCGCTGTGCGCAATGTAACACTTTGACAACTTGACAACTGAGT 32099
Qy 181 tdcgtgaggggtcttccctcgtcagtgaggatttaccgagattcaggaatgggttttctccc 240
Db 32098 TAGTCCGAGGCTTCCCTGCACTGTGGGATTTACGCTGATTCAGGAATGGGTTGTCTCC 32039
Qy 241 ttggatattgttcttaacgcgggagagcttgaatccctcgtgaggaagtgtatgcacgtgac 300
Db 32038 TGGATATGTTCTTAACCGGGGAGGAGCTTGAATCCTCAGGAAGTGTATGCACGTGTGC 31979
Qy 301 ctgtgttgcccaacttgcataatgcagcagcagatgcacatggtttacagagtcctgg 360
Db 31978 CTGTGTTGTGCCAATGATATCATGAGAGCATGATCATGTTTACGAGTCTCTGG 31919
Qy 361 gctctccactgtcattgttccagctccggttccctcagtgatagcccgcgggcgaggtt 420
Db 31918 GCTCTCCACTGTCTATTGTTCCAGTCCCGGTTCCCTCAGTGTATATGCCGGGGCAGGTT 31859
Qy 421 ttggccagctggtttagagtggtggtggatggcgccatgtttaatcagaggttttatctg 480
Db 31858 TTGGCCAGCTGTTAGGATGTTGGTGGATGGCGCCATGTTTAAATCAGAGGTTTATATGG 31799
Qy 481 tacgggaggtgtgaattacaacatgcacaaagagagtaattgtttatccacgctggtt 540
Db 31798 TACGGGAGGTGTGTAATTAACAATGCCAAAGAGAGTAATGTTTATGTCACACGTTGTT 31739
Qy 541 atgaggggtgcaccttaattcactcgtcgttgggtatgattgagcagcagtggttctctg 600
Db 31738 ATGAGGGGTGCCACTTAATCTACCTCGCTGTGGTATGATGCCACGTTGGTCTGTG 31679
Qy 601 gtccccccatgagcttggatgacagcgcccttgcaactgtgggattttgacaacattgtg 660
Db 31678 GTCCCGCCATGAGCTTTGGATACAGCGCCTTGCACTGTGGGATTTTGAACAATATTGTG 31619
Qy 661 gtgtgtgtcagcttactgtctgatttaagtgtgagatcaggggtgctgtgtgccccgg 720
Db 31618 GTGTGTGCTGCAGTTACTGTGCTGATTTAAGTGTGAGATCAGGGTGCCTGCTGTGCCCG 31559
Qy 721 aggcacagcgcccttattctcgtcggcggtgcgaatcatcgtgagggagacactgcccattg 780
Db 31558 AGGACAGCGCCITATGCTCGGCGGTGCGAATCATCGCTGAGAGACACTGCCATG 31499
Qy 781 ttgtattcttcagagcagcgagcgcggtggcgagcagtttattcgcgctgctcagcac 840
Db 31498 TTGTATTCTTCAGGACGAGCGCGCGGCGGAGGATTTATTCGCGCGCTGCTGCAGCAC 31439
Qy 841 caccgcccctactctgacgatttactgacttaccctcccatgtag 885
Db 31438 CACCGCCCTATCCTGATGACGATTTATGACTCTACCCCATGTAG 31394

RESULT 4
US-08-735-609-1/c
; Sequence 1, Application US/08735609
; Patent No. 5955360
; GENERAL INFORMATION:
; APPLICANT: Chamberlain, Jeffrey S.
; APPLICANT: Amalfitano, Andrea
; APPLICANT: Hauser, Michael A.
; APPLICANT: Kumar-Singh, Rajendra
; APPLICANT: Hartigan-O'Connor, Dennis J.
; TITLE OF INVENTION: IMPROVED ADENOVIRUS VECTORS
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll, LLP

```
Query Match      100.0%; Score 885; DB 5; Length 35935;
Best Local Similarity 100.0%; Pred. No. 7.2e-282;
```

Matches 885; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgactacgctccgcttcattgcatgacactacgacacacgactcgggtgtct 60
DB 34077 ATGACTACGCTCCGCGCTTCATTGGCTGACACTACGACCAACAGATCGGTGTCT 34018

QY 61 cggcgactccgtacagtagggatcgtctacotccttttgagacagaaacccgcgtacc 120
DB 34017 CGGCGCACTCCGTACAGTAGGATCGTCTACCTCCCTTTGAGACAGAAACCCGCGTACC 33958

QY 121 atactgagatcatccgctgctccgcaatgtaacacttgacaatgcacacgtagt 180
DB 33957 ATACTGAGGATCATCCGCTGCTCCGCAATGTAACACTTTGACAAATGCACACGAGT 33898

QY 181 tacgtcgaggtctccctgcagtggtggatttacgtgattcaggaatgggtgtctcc 240
DB 33897 TACGTGGAGGTCTCCCTGCAGTGGGATTTACGCTGATTGAGGAATGGGTGTCTCC 33838

QY 241 cgggataggtcttaacgcgagggagccttgtaactcctgaggaagtatgcagctgtgc 300
DB 33837 TGGGATATGTTCTAACCGCGGAGAGCTTGTAATCTGAGGAAGTGTATGCAGGTGTC 33778

QY 301 ctgtgtgtgccaacatgatcatcacgagcagatgatccatggtttacgagtcctgg 360
DB 33777 CTGTGTGTGCGCAACATTCATATCATGACGAGCATGATCCATGTTACGAGTCTGG 33718

QY 361 gctccactgtcattgttccagtcocgggttccctgcagtgatagcggcgagcagtt 420
DB 33717 GCTCTCCACTGTCTATGTTCCAGTCCCGGTTCCCTGAGTGTATAGCGCGCGCAGTT 33658

QY 421 ttggccagctggttagataggtgtgtagtgcgacatgttttaacagaggtttatgg 480
DB 33657 TTGGCCAGCTGGTTTAGGATGTGTGTGAGTGGCGCATGTTTATCAGAGTTTATGG 33598

QY 481 taccggaggtgtgtaatacaacatgcacacagaggaaggttaattgttccagcgtgtt 540
DB 33597 TACCGGAGGTGTGTAATTAACAACATGCCAAAGAGGTAATGTTATGTCAGGCGTGT 33538

QY 541 atgaggggtgcacatcaatcaactacgtgctgtgtagtgcagcagtggtgtctgtg 600
DB 33537 ATGAGGGGTGCGCACATTAATCTACCTGCGCTGTGTGATGATGCGCATGTTGTTG 33478

QY 601 gtcccgccatagcttctgacacagcgtctgcactgtgggattttgaacaattgtg 660
DB 33477 GTCCCGCCATGAGCTTTGGATACAGCGCTTGACATGCGATTTGAACAATATTGTG 33418

QY 661 gtgctgtgtcagttactgtgtgttgaatgaatgaatgaatgaatgaatgaatgaat 720
DB 33417 GTGCTGTGTGCAAGTACTGTGTGATTTAAGTGATCAGGCTGCTGTGTGCTGCTG 33358

QY 721 aggaacagggccttattctgctggcggtgcgaatcatcgtgagagaccactgccatg 780
DB 33357 AGGACAGGCGCTTATGCTGCGGCGGTGCGNATCATGCTGAGGAGACCACTGCCATG 33298

QY 781 ttgtattcctgcagcagcagcggcgcgagcagctttattogcgtgtgtgcagcac 840
DB 33297 TTGTATTCTGACAGAGCGGAGCGCGGCGCAGCAGTATTATTCGCGCTGTGTGACGAC 33238

QY 841 caccgcctctctgtatgcagctattatgactctacccccatgtag 885
DB 33237 CACGCGCTATCTGTATGTCAGTATTATGACTTACCCCCATGTAG 33193

RESULT 6
US-08-752-760A-1/c
; Sequence 1, Application US/08752760A
; Patent No. 5877011
; GENERAL INFORMATION:
; APPLICANT: Armentano, Donna
; APPLICANT: Gregory, Richard J.
; APPLICANT: Smith, Alan E.
; TITLE OF INVENTION: CHIMERIC ADENOVIRAL VECTORS
; NUMBER OF SEQUENCES: 3

CORRESPONDENCE ADDRESS:
ADDRESSEE: Baker & Botts, L.L.P.
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: NY
COUNTRY: U.S.A.
ZIP: 10112

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08752,760A
FILING DATE: 20-NOV-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Seide, Rochelle K
REGISTRATION NUMBER: 32,300
REFERENCE/DOCKET NUMBER: A31385
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-705-5000
TELEFAX: 212-705-5020
TELEX:

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 35081 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-752-760A-1

Query Match 36.1%; Score 319.2; DB 3; Length 35081;
Best Local Similarity 62.4%; Pred. No. 5.7e-95;
Matches 520; Conservative 0; Mismatches 308; Indels 6; Gaps 1;

QY 46 cgatctcgtgtgtctcggcgacactccgtacagtagggatcgtctacacctcttttgagaca 105
DB 33153 CAATCCAGCTCGCTCCGCCATCATCATATCGCAGGCGCGTTCACCAAGATCTGATGAT 33094

QY 106 gaaacccgcgtaccactactcggagggatccctcctcctcagtggtggatttadgctgacag 165
DB 33093 GAAACCCGCGCTTCACTGACTGATCAACATCCCTCTGTGCCCGATGTGTGTCATGCGAT 33034

QY 166 atgcacacgtgagttacgtgcaggtcttccctcagtggtggatttadgctgacag 225
DB 33033 TATCATTAATGTAAGTTCTGTCGCGGATACCATGTGCTGCTGCTGCTGCTGCTGCTGCT 32974

QY 226 gaatgggtgttccctcggataggttctaacgcgggagagcttgaatccctgaggaag 285
DB 32973 GAGTTTCAGTCCCTCGGATATGATCCTCACCACAGAGGAAATATAAAATTTTAAAGA 32914

QY 286 tgtatgcagctgtcctcgtgtgttgcccaactgatcatcagcagcagcatgatccat 345
DB 32913 TGTATGTCAGTGTGCTGTCGCCCGCTACCTGAGCTTGGTGAAGAGCTCAGATGGTGAGC 32854

QY 346 ggttacgagctcgtggctcctcactgtcattgttccagtcctcccggttccctcagtgata 405
DB 32853 GGTACAGAGCGCTGGATCTCATTGTCACATGTCACATGTCACCTCCCTGCTGCTGCTGCTG 32794

QY 406 gccggcgaggtttttggccagctggttttaggagtggtgggtgggtgggtgggtgggtgggt 465
DB 32793 GCGGAGGACACCTGCTGCGCGTGTGTGTGAGAGAGTCAATTTACGGGTGATGTTCAAC 32734

QY 466 cagaggtttataggtaccgaggggtggtgaattaccacacagcgaagaggttaatttt 525
DB 32733 CAGCGCTTCCCTGTCACCGCAGATTGTAACAGAAACATGCCCAAGAGATCATGTAT 32674

QY 526 atgtccagcgtgtttatgaggggtgcgccacttaattctacctgcgcttgggtatgatggcg 585

	Query Match	3.5%;	Score 30.6;	DB 2;	Length 7336;
	Best Local Similarity	53.8%;	Pred. No. 5.2;		
	Matches 63;	Conservative 0;	Mismatches 54;	Indels 0;	Gaps 0;
QY 628	gccttgcaactgtgggattttgaacaatatattggtgtgtctgctgcagttactgtctgat	687			
Db 6296	GCCATTAGCTGTGGTGTGATTAAAGATGTTGACGACCTGCTGCTGCTGAGGAGAG	6237			
QY 688	ttaagtgaatcaggggtgcgtgctgtgcccgaggacaaaggcgccctatgctgcgg	744			
Db 6236	CTGGGTGACATTTAGGGTCCCTAGCAGCTCTTTGACGGSCACTTTGCGAGATGCTCTG	6180			

RESULT 11
US-08-348-891A-1
; Sequence 1, Application US/08348891A
; Patent No. 5654136
; GENERAL INFORMATION:
; APPLICANT: SASAKI, Keiko
; APPLICANT: MORI, Takayuki
; APPLICANT: MAKINO, Satoshi
; TITLE OF INVENTION: ATTENUATED MEASLES VIRUS VACCINE,
; CONTAINING SPECIFIC NUCLEOTIDE SEQUENCE AND A METHOD FOR
; ITS ABSOLUTE IDENTIFICATION
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: YOUNG & THOMPSON
; STREET: 745 South 23rd Street
; CITY: Arlington
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/348,891A
; FILING DATE: 25-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/848,400
; FILING DATE: 10-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 3-293625
; FILING DATE: 14-OCT-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: PATCH, Andrew J.
; REGISTRATION NUMBER: 32,925
; REFERENCE/DOCKET NUMBER: KP-7501
; TELEPHONE: 703-521-2297
; TELEFAX: 703-685-0573
; TELEX: 248425 EMBON
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15894 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 108..1682
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1807..3327
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3438..4442
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 5458..7107
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 7271..9121
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 9234..15782
; US-08-348-891A-1
Query Match 3.5%; Score 30.6; DB 1; Length 15894;
Best Local Similarity 49.1%; Pred. No. 8;

Matches 81; Conservative 0; Mismatches 84; Indels 0; Gaps 0;
QY 91 cctccttttgagacagaaacccgcgtaccatactggagagatcatcctgctgccccgaa 150
|||||
Db 2516 CCTCATTTGGAACGGAGATCGCGTCTTTATTGACAGGTGGTGCACCAATGTGCTCGAA 2575
QY 151 tgtaaacctttgacaatgcacaacgtgagttacgtgagaggtcttccctcagtgaggga 210
|||||
Db 2576 AGTCACCTCGGAACCAATCAGGGCCAGTGCACCTCGGGGAATGTCCCGAGTATGTGA 2635
QY 211 tttaacgtgattcagggaatgggtttccctgggatggttcta 255
|||||
Db 2636 GCAATGCCGACATGATACAGAGTGGACACCCGGAATCTGTGTACCA 2680
RESULT 12
US-08-905-817-1
; Sequence 1, Application US/08905817
; Patent No. 5824777
; GENERAL INFORMATION:
; APPLICANT: SASAKI, Keiko
; APPLICANT: MORI, Takayuki
; APPLICANT: MAKINO, Satoshi
; TITLE OF INVENTION: ATTENUATED MEASLES VIRUS VACCINE,
; CONTAINING SPECIFIC NUCLEOTIDE SEQUENCE AND A METHOD FOR
; ITS ABSOLUTE IDENTIFICATION
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: YOUNG & THOMPSON
; STREET: 745 South 23rd Street
; CITY: Arlington
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/905,817
; FILING DATE: 04-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/348,891
; FILING DATE: 25-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/848,400
; FILING DATE: 10-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 3-293625
; FILING DATE: 14-OCT-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: PATCH, Andrew J.
; REGISTRATION NUMBER: 32,925
; REFERENCE/DOCKET NUMBER: KP-7501A
; TELEPHONE: 703-521-2297
; TELEFAX: 703-685-0573
; TELEX: 248425 EMBON
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15894 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 108..1682
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1807..3327

Qy 319 gatatacagcagcatgatcatcattggtctccactgctcattgt 378

Search completed: June 23, 2000, 09:51:07
Job time: 6653 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run on: June 23, 2000, 05:42:32 ; Search time 592.38 Seconds
(without alignments)
6055.408 Million cell updates/sec

Title: US-09-214-478-1
Perfect score: 885
Sequence: 1 atgactacgtcgcgcgttcc.....atgactatcccccatgtag 885

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues 9714632
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 1000000
Post-processing: Minimum Match 0%
Listing first 45 summaries

Database:

- 1: em_est1:*
- 2: em_est2:*
- 3: em_est3:*
- 4: em_est4:*
- 5: em_est5:*
- 6: em_est6:*
- 7: em_est7:*
- 8: em_est8:*
- 9: em_est9:*
- 10: em_est10:*
- 11: em_est11:*
- 12: em_est12:*
- 13: em_est13:*
- 14: em_est14:*
- 15: em_est15:*
- 16: em_est16:*
- 17: em_est17:*
- 18: em_est18:*
- 19: em_est19:*
- 20: gb_est1:*
- 21: gb_est2:*
- 22: gb_est3:*
- 23: gb_est4:*
- 24: gb_est5:*
- 25: gb_est6:*
- 26: gb_est7:*
- 27: gb_est8:*
- 28: gb_est9:*
- 29: gb_est10:*
- 30: gb_est11:*
- 31: gb_est12:*
- 32: gb_est13:*
- 33: gb_est14:*
- 34: gb_est15:*
- 35: gb_est16:*
- 36: gb_est17:*
- 37: gb_est18:*
- 38: gb_est19:*
- 39: gb_est20:*
- 40: gb_est21:*
- 41: gb_est22:*
- 42: gb_est23:*
- 43: gb_est24:*
- 44: gb_est25:*

- 45: gb_est26:*
- 46: gb_est27:*
- 47: gb_est28:*
- 48: gb_est29:*
- 49: gb_est30:*
- 50: gb_est31:*
- 51: gb_est32:*
- 52: em_est20:*
- 53: em_est21:*
- 54: em_est22:*
- 55: em_est23:*
- 56: em_est24:*
- 57: em_est25:*
- 58: em_est26:*
- 59: gb_est33:*
- 60: gb_est34:*
- 61: gb_est35:*
- 62: gb_est36:*
- 63: gb_est37:*
- 64: gb_est38:*
- 65: em_est27:*
- 66: em_est28:*
- 67: em_est29:*
- 68: em_est30:*
- 69: gb_est39:*
- 70: gb_est40:*
- 71: gb_est41:*
- 72: gb_est42:*
- 73: gb_est43:*
- 74: gb_est44:*
- 75: em_est31:*
- 76: em_est32:*
- 77: em_est33:*
- 78: em_est34:*
- 79: gb_est45:*
- 80: gb_est46:*
- 81: gb_est47:*
- 82: gb_gss1:*
- 83: gb_gss2:*
- 84: gb_gss3:*
- 85: gb_gss4:*
- 86: em_gss1:*
- 87: em_gss2:*
- 88: em_gss3:*
- 89: em_gss4:*
- 90: gb_gss5:*
- 91: gb_gss6:*
- 92: gb_gss7:*
- 93: gb_gss8:*
- 94: gb_gss9:*
- 95: em_gss5:*
- 96: em_gss6:*
- 97: em_gss7:*
- 98: em_gss8:*
- 99: em_gss9:*
- 100: em_gss10:*
- 101: em_gss11:*
- 102: gb_gss10:*
- 103: gb_gss11:*
- 104: em_gss12:*
- 105: gb_gss12:*
- 106: gb_gss13:*
- 107: gb_gss14:*
- 108: gb_gss15:*
- 109: gb_gss16:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result Query

No.	Score	Match	Length	DB	ID	Description
1	521	58.9	564	79	AW265091	xp79d11.x
2	503	612	79	AW277220	AW277220	xq9b06.x
3	228.6	25.8	515	79	AW277227	xq9b06.x
4	228.6	25.8	515	81	AW419027	xy45f04.x
5	228.6	25.8	516	81	AW438779	xy38e04.x
6	218.8	24.7	493	80	AW340909	xy50e09.x
7	213.2	24.1	429	36	AA652034	ns48h01.s
8	201.4	22.8	301	36	AA649817	ns48f01.s
9	195.8	22.1	223	79	AW270876	xq56b11.x
10	194.6	22.0	445	36	AA642819	nu5B06.s
11	43.8	4.9	630	79	AW276836	xp66g11.x
12	42.2	4.8	922	82	CNS0073W	AW276836
13	40.4	4.6	848	82	CNS00GQW	AW276836
14	40.2	4.5	393	109	AO640309	AW276836
15	39.2	4.4	1101	82	CNS00L2R	AW276836
16	38.8	4.4	1101	82	CNS00K22	AW276836
17	38.4	4.3	1101	83	CNS0161W	AW276836
18	38	4.3	895	82	CNS0071A	AW276836
19	37	4.2	424	109	AO648430	AW276836
20	37	4.2	537	92	AO944567	AW276836
21	37	4.2	544	92	AO940148	AW276836
22	36.6	4.1	511	91	AO851204	AW276836
23	36	4.1	666	47	AT531303	AW276836
24	35.4	4.0	939	82	CNS00CNG	AW276836
25	35.2	4.0	681	91	AO899964	AW276836
26	35.2	4.0	827	84	B12606	AW276836
27	35.2	4.0	1101	83	CNS016HE	AW276836
28	34.8	3.9	434	48	AT596460	AW276836
29	34.8	3.9	442	109	AO647993	AW276836
30	34.8	3.9	595	92	AO919044	AW276836
31	34.8	3.9	902	82	CNS006QP	AW276836
32	34.6	3.9	380	41	AI019207	AW276836
33	34.4	3.9	1101	82	CNS0038X	AW276836
34	34.2	3.9	400	31	AA294303	AW276836
35	34	3.8	600	106	AO369543	AW276836
36	34	3.8	791	60	AT806482	AW276836
37	34	3.8	1101	82	CNS00B8O	AW276836
38	33.6	3.8	339	22	R72680	AW276836
39	33.6	3.8	415	33	AA402831	AW276836
40	33.6	3.8	429	34	AA459525	AW276836
41	33.6	3.8	1101	83	CNS017ZT	AW276836
42	33.4	3.8	511	23	R30064	AW276836
43	33.4	3.8	536	63	AI994566	AW276836
44	33.2	3.8	399	49	AI660346	AW276836
45	33.2	3.8	436	61	AI874404	AW276836

ALIGNMENTS

RESULT 1
 LOCUS xp79d11.x1 NCI-CGAP_Ov40 Homo sapiens cDNA clone IMAGE:2746581 3'
 DEFINITION similar to SW:E434_ADE02 P03239 EARLY E4 34 KD PROTEIN. ;, mRNA
 EST 28-DEC-1999
 accession AW265091 564 bp mRNA
 version AW265091.1 GI:5641907
 keywords EST.
 source human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 564)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 On Jul 7, 1999 this sequence version replaced gi:5406367.
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov
 Tissue Procurement: Andrew Berchuck M.D., John Gillespie M.D.,
 Michael R. Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: David B. Krizman, Ph.D.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www-bio.llnl.gov/bbrp/image/image.html

Possible reversed clone: polyt not found
 Seq primer: -40UP from Gibco
 High quality sequence stop: 438.

FEATURES

Location/Qualifiers
 1..564
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2746581"
 /clone_lib="NCI-CGAP_Ov40"
 /sex="female"
 /tissue_type="endometrioid ovarian metastasis"
 /lab_host="DH10B"
 /note="Organ: ovary; Vector: pAMP10; cDNA made by oligo-dT
 priming. Non-directionally cloned into the UDG sites of
 pAMP10. Size-selected on agarose gel, average insert
 size 500 bp. Primary library, non-amplified. CDNA
 Library Preparation: David B. Krizman, Ph.D (NCI).
 Reference: Krizman et al. (1996) Cancer Research
 56:5380-5383."

BASE COUNT 158 a 175 c 117 g 112 t 2 others
 ORIGIN

Query Match 58.9%; Score 521; DB 79; Length 564;
 Best Local Similarity 98.7%; Pred. No. 3.3e-147;
 Matches 524; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 121 atactgagagatcatccgctgctgcccgaatgaacattgacaatgcaacagtgagt 180
 Db 564 ATACTGNAGGATCATCGCTGCTGCCGAATGTAACACTTTGACATGCAACAGTGAGT 505
 QY 181 taagtgcaggttccctcagtggtggatttacgtgtattcaggaaatgggtgtcc 240
 Db 504 TACGTGCGAGGCTTCCCTGAGTGGGATTACGCTGATTCAGGAATGGGTGTCCN 445
 QY 241 tgggatatgttctaacgcggaggagcttgtaactcctgagggaagtgtatgcacgtgtgc 300
 Db 444 TGGGATATGTTCTGACGCGGAGGAGCTTGTAACTCCTGAGGAAGTGTATGCACGTGTGC 385
 QY 301 ctgtgttgcccaacattgatcatcagcagcatgatccatggttacgagtcctgg 360
 Db 384 CTGTGTTGTGCCAACATTGATATCATGACGAGCATGATCATCCATGTTACGAGTCTGG 325
 QY 361 gctctccatgtcatgttccagtcgccggttccctcagtcagtgatagccggcgagggtt 420
 Db 324 GCTCTCCACTGTCTATTGTTCCAGTCCCGTTCCTGAGTGCATAGCCGCGGAGGTT 265
 QY 421 ttggccagctgggtttaggatgggtggatggcgcgaatttaatacagaggtttatagg 480
 Db 264 TTGGCCAGCTGTTTGGATGTTGGTGGATGGCGCCATGTTTAAACAGAGGTTTATATGG 205
 QY 481 taccggaggtggtgaattacacatgccaaagaggttaattttatgtccagcgtgttt 540
 Db 204 TACCGGAGGTGGTGAATTAACACATGCCAAAGAGGTAATGTTTATGTCACGCGTGT 145
 QY 541 atgaggggtcgcacctaactacgtggtgtgtgtatgagccacgtgggttctgtg 600
 Db 144 ATGAGGGGTGCGCACCTAATCTACCTGCGCTTGTGGTATGATGCGCACGTGGGTCTGTG 85
 QY 601 gtccccgccatgagctttggatgacagcgccttgactgtgggattttgaac 651
 Db 84 GTCCCCGCCATGAGCTTTGGATGACAGCGCCITGCACTGTGGATTTTGAAC 34

[illegible]

RESULT	11
AWAV275636	
LOCUS	
DEFINITION	
ACCESSION	
VERSION	
KEYWORDS	
SOURCE	
ORGANISM	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
COMMENT	
FEATURES	
SOURCE	
BASE COUNN	
ORIGIN	
Query M	
Best Loc	
Matches	
QY	1
Db	584
RESULT	1
CNS0073W	
LOCUS	
DEFINITION	
ACCESSION	
VERSION	

... ..

AL072642
VERSION
AL072642.1 GI:4952523
GSS.
SOURCE
fruit fly,
Drosophila melanogaster
ORGANISM
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 848)
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequen-
- BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.c
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part
collaboration with the Berkeley Drosophila Genome Project (BDG)
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammosser in Pieter de Jong's laboratory in the Department
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from three
isogenic strain y2; cn bw sp, the same strain used for the BDGP
p1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
source 1. .848
Location/Qualifiers
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

Query Match	4.6%;	Score 40.4;	DB 82;	Length 848;
Best Local Similarity	22.6%;	Pred. NO. 0.18;		

Query Match	4.6%	Score 40.4;	DB 82;	Length 848;
Best Local Similarity	22.6%;	Pred. No. 0.18;		
Matches 42;	Conservative 74;	Mismatches 70;	Indels 0;	Gaps
375	ti g tccaggtcccggttccctgcagtgatagccggcgaggttttggccagctggtt	434		
	: : : : : : : : : :			
703	KT K YSB G K B TGKKKKGTGKKGGDGGKGVKSSGGGGGGKGTGGGGKCKGKK	644		
	: : : : : : : : :			
435	taggagtggtggtagtggcgcatgtttaatcacagaggttatgatgcaggagagtg	494		
	: : : : : : : : :			
643	GNDGGKKKGGARGGTTTTTKTKTKKTAMDWA A AAAKKATDTKKKGTCKRAGTGGK	584		
	: : : : : : : : :			
495	gaattacaaatgcga a aaagagtaagtattatgcagcgtgtttatgaggggtgc	554		
	: : : : : : : : :			
583	RA R ADGKAGKADGAADKKRKGKWTATTTT T KKAAWKGS T TTTAAWKT T KDKWA	524		
	: : : : : : : : :			
555	cttaat 560			
	: : : : : : : : :			
523	TW W AK 518			

RESULT 14	393 bp	DNA	GSS	08-JUL-1999
0640309	927P1-4G2.TP	Trypanosoma brucei	genomic clone 927P1-4G2,	
CCUS	genomic survey sequence.			
DEFINITION	AQ640309			
ACCESSION	AQ640309			
VERSION	AQ640309.1			
KEYWORDS	GSS.			
SOURCE	Trypanosoma brucei.			
ORGANISM	Trypanosoma brucei			
	Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;			
	Trypanosoma.			

REFERENCE	1 (bases 1 to 393)
AUTHORS	El-Sayed,N., Zhao,S., Zhao,H., Gill,S.I., Suh,E., Malek,J., Fujii,C., Gerrard,C., Leech,V., de Jong,P., Ullu,E., Melville,S., Donaldson,J., Fraser,C. and Adams,M.
TITLE	Determination of clone end sequences from Trypanosoma brucei TREU 927/4 P1 library
JOURNAL	Unpublished (1999)
COMMENT	On Mar 23, 1999 this sequence version replaced g1:3323624. Other_GSSs: 927P1-4G2.TV Contact: Najib M. El-Sayed Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200 Fax: 301 838 0208 Email: nelsayed@tigr.org For clone/filter availability, please contact Sara Melville (sm160@molc.bio.cam.ac.uk). P1 end sequences search page: http://www.tigr.org/tdb/mdb/tbdb/ . Seq primer: SP6 Class: P1 ends.

FEATURES	source
Class: 1	Location/Qualifiers
1. .393	
/organism="Trypanosoma brucei"	
/strain="TREU927/4"	
/db_xref="taxon:5691"	
/clone="927Pl-4c2"	
/clone_lib="927Pl"	
/note="Vector: pAD10SacBII; Site_1: Bam HI; Constructed by Sara Melville, University of Cambridge, UK and Nancy Shepherd, Dupont Merck, Wilmington, DE. Genomic DNA was isolated from Trypanosoma brucei (stock TREU927/4) and partially digested with Sau 3AI. DNA fragments were cloned into the Bam HI site of pAD10SacBII vector (Genbank accession U09128). The average insert size is 65 Kb. Coverage: approx 4.4 x the haploid non-minichromosomal genome."	
80 a	87 c 134 g 92 t
BASE COUNT	
ORIGIN	

	Query Match	4.5%	Score 40.2;	DB 109;	Length 393;
	Best Local Similarity	54.4%;	Pred. No. 0.15;		
	Matches 381;	Conservative	0;	Mismatches	68;
				Indels	0;
				Gaps	0;
QY	500	aacacatgccaaaagaggtaatgtttatgtccagcgtgtttatgaggggtgcgccctaa	559		
DB	195	ACGGGATGCCGATGCGGCTTCGTTTAAATGATGATGCCGTGAACCTGCGACACTGC	254		
QY	560	tctactcgcgtgtgggatgatgagccacgtgggtctctgtgtcccccgcacatgagctttg	619		
DB	255	GCTGTGAGGCGGCATGTAGGGCGTGACCGCTGACGCATGCGAATTCGCCGATGAGATGTA	314		
QY	620	gatacagcgccttgcactgtgggattttg	648		
DB	315	CATACACACACTGGAGCTCTGTGACATFG	343		

RESULT	15
CNS00LT2	DNA
LOCUS	1101 bp GSS 14-JUN-1999
DEFINITION	Drosophila melanogaster genome survey sequence TET3 end of BAC:
	BACH48P19 of RPCI-98 library from Drosophila melanogaster (fruit
	fly), genomic survey sequence.
ACCESSION	AL078714
VERSION	AL078714.1 GI:5102004
KEYWORDS	GSS.
SOURCE	fruit fly.
ORGANISM	Drosophila melanogaster
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE	1 (bases 1 to 1101)

AUTHORS
TITLE Genoscope.
JOURNAL Direct Submission
Submitted (11-JUN-1999) Genoscope - Centre National de Sequençage
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
- Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammeter at the Roswell Park Cancer Institute in Buffalo NY. The library is named RP11-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp. The same strain used for the library P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at <http://bacpac.mecb.buffalo.edu/drosophila.bac.htm>

[illegible]

Search completed: June 23, 2000, 09:17:09
Job time: 12877 sec

Mon Jun 26 09:10:32 2000

us-09-214-478-1.rst

Page 10

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM.nucleic - nucleic search, using sw model

Run on: June 23, 2000, 09:34:53 ; Search time 878.41 Seconds

(without alignments)
-382.069 Million cell updates/sec

Title: US-09-214-478-3
Perfect score: 345
Sequence: 1 atgtttcttcagcttttc.....cggcgactcgtacagtag 345

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 1765538

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

GenEmbl.*

- 1: gb_ba1.*
- 2: gb_ba2.*
- 3: gb_cm.*
- 4: gb_ov.*
- 5: gb_pat.*
- 6: gb_ph.*
- 7: gb_pl1.*
- 8: gb_pl2.*
- 9: gb_pr1.*
- 10: gb_pr2.*
- 11: gb_pr3.*
- 12: gb_ro.*
- 13: gb_sts.*
- 14: gb_sy.*
- 15: gb_un.*
- 16: gb_vi.*
- 17: em_fun.*
- 18: em_hum1.*
- 19: em_hum2.*
- 20: em_in.*
- 21: em_om.*
- 22: em_or.*
- 23: em_ov.*
- 24: em_pat.*
- 25: em_ph.*
- 26: em_p.*
- 27: em_ro.*
- 28: em_sts.*
- 29: em_sy.*
- 30: em_un.*
- 31: em_vi.*
- 32: gb_htg1.*
- 33: gb_htg2.*
- 34: gb_in1.*
- 35: gb_in2.*
- 36: gb_ba1.*
- 37: gb_ba2.*
- 38: em_hum3.*
- 39: em_hum4.*
- 40: gb_pr4.*
- 41: gb_htg3.*
- 42: gb_htg4.*
- 43: gb_htg5.*
- 44: gb_htg6.*

- 45: gb_htg7.*
- 46: em_htg1.*
- 47: em_htg2.*
- 48: em_htg3.*
- 49: em_hum5.*
- 50: gb_pl3.*
- 51: gb_pr5.*
- 52: gb_htg8.*
- 53: gb_htg9.*
- 54: gb_htg10.*
- 55: gb_htg11.*
- 56: gb_htg12.*
- 57: gb_htg13.*
- 58: gb_htg14.*

Complement

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
C 1	345	100.0	3189	5	A52460	A52460 Sequence 4
C 2	345	100.0	35935	16	ADRCMPGEN	M73260 Mastadenovirus
C 3	341.8	99.1	35937	16	ADRCG	J01917 Adenovirus
C 4	80	23.2	8299	5	AR009152	AR009152 Sequence
C 5	54	15.7	3119	16	S82508	S82508 984: Orf2..
C 6	54	15.7	35100	16	AF108105	AF108105 Human ade
C 7	52.8	15.3	3498	16	ADERB4	AF1800 Adenovirus
C 8	52.8	15.3	34125	16	AT12CGA	X73487 Adenovirus
C 9	36.8	10.7	34214	16	ADRGENOME	L19443 Human adenov
C 10	36.2	10.5	170245	10	HS109F14	AL022721 Human DNA
C 11	36	10.4	9971	12	NMGFPAD	X02801 Mouse gene
C 12	35.2	10.2	773	9	HUMSTXG1	M23945 Human chrom
C 13	35.2	10.2	2401	9	HUMSTB	J04964 Human stero
C 14	35.2	10.2	6520	9	HUMSTB	M15505 Human stero
C 15	35.2	10.2	99033	11	AC005704	AC005704 Homo sapi
C 16	35	10.1	486	3	AF075663	AF075663 Equus cab
C 17	35	10.1	14634	2	AE001442	AE001442 Helicobac
C 18	33.8	9.8	45410	41	AC006187	AC006187 Homo sapi
C 19	33.8	9.8	83336	55	AC022747	AC022747 Homo sapi
C 20	33.6	9.7	112629	8	F28D6	AF147252 Arabidops
C 21	33.4	9.7	211150	45	AC008732	AC008732 Homo sapi
C 22	33.2	9.6	6448	1	SCU68036	U68036 Streptomyce
C 23	32.6	9.4	214520	10	CNS0000J	AL049839 Human chr
C 24	32.6	9.4	278334	56	AC008543	AC008543 Homo sapi
C 25	32.4	9.4	40582	9	AP000691	AP000691 Homo sapi
C 26	32.2	9.3	872	40	AF112206	AF112206 Homo sapi
C 27	32.2	9.3	166180	32	CNS01DRF	AL117190 Homo sapi
C 28	32.2	9.3	184969	44	AC018362	AC018362 Homo sapi
C 29	32	9.3	4205	1	SGJ000264	AJ000264 Streptomy
C 30	31.8	9.2	771	13	G41969	G41969 SHGC-68997
C 31	31.8	9.2	35377	1	MTCY48	274020 Mycobacteri
C 32	31.8	9.2	112456	51	AC004614	AC004614 Homo sapi
C 33	31.8	9.2	216184	45	AC015899	AC015899 Homo sapi
C 34	31.8	9.2	225432	11	AF027390	AF027390 Homo sapi
C 35	31.6	9.2	188267	41	AC007734	AC007734 Homo sapi
C 36	31.4	9.1	1173	50	AF190094	AF190094 Magnolia
C 37	31.2	9.0	130690	1	D90901	D90901 Synchocyst
C 38	31	9.0	18746	42	AC015350	AC015350 Drosophill
C 39	31	9.0	118235	41	AC007854	AC007854 Drosophill
C 40	31	9.0	166536	56	AC008772	AC008772 Homo sapi
C 41	31	9.0	179127	41	AC007378	AC007378 Homo sapi
C 42	30.8	8.9	1735	11	HSU48361	U48361 Human NGFI-
C 43	30.8	8.9	1987	9	HS011081	AJ011081 Homo sapi
C 44	30.8	8.9	2192	10	HSMADR	X70991 H sapiens M
C 45	30.8	8.9	35698	2	AF213822	AF213822 Zymomonas

ALIGNMENTS

1. RESULT 1
LOCUS A52460 3189 bp DNA PAT 12-DEC-1997
DEFINITION Sequence 4 from Patent WO9622378.
ACCESSION A52460
VERSION A52460.1 GI:2851930
KEYWORDS
SOURCE unclassified.
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 3189)
AUTHORS Dedieu J., Latta M., Orsini C., Perricaudet M., Vigne E. and Yeh P.
TITLE CELLS FOR THE PRODUCTION OF RECOMBINANT ADENOVIRUSES
JOURNAL PATENT: WO 9622378-A 4 25-JUL-1996;
COMMENT RHONE-POULENC RORER SA (FR)
Other publication AU 4544396 960807
Other publication FR 2729674 960726.
FEATURES
Location/Qualifiers
1..3189
/organism="unclassified"
/db_xref="taxon:32644" 671 t
BASE-COUNT 975 a 912 c 631 g
ORIGIN
Query Match 100.0%; Score 345; DB 5; Length 3189;
Best Local Similarity 100.0%; Pred. No. 5.5e-97;
Matches 345; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 atgggtctccagctctccgctccctccgctgtgtgactgcagacgaatgtgaggt 60
DB 1593 atgggtctccagctctccgctccctccgctgtgtgactgcagacgaatgtgaggt 1534
QY 61 tggctgggtggtggtattctgcggtggtggtggtggtggtggtggtggtggtggt 120
DB 1533 tggctgggtggtggtattctgcggtggtggtggtggtggtggtggtggtggtggt 1474
QY 121 gttacatagacacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 180
DB 1473 gttacatagacacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 1414
QY 181 aactactacagacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 240
DB 1413 aactactacagacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 1354
QY 241 cgcacgtggttcttcaggaatagactacgctcgcggttcatttgcgacgacgac 300
DB 1353 cgcacgtggttcttcaggaatagactacgctcgcggttcatttgcgacgacgac 1294
QY 301 acgacacacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 345
DB 1293 acgacacacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 1249
2. RESULT 2
LOCUS ADRCOMPEN/35935 bp DNA VRL 08-APR-1996
DEFINITION Mastadenovirus h5 gene, complete genome.
ACCESSION M73260 M29978
VERSION M73260.1 GI:209842
KEYWORDS protein 5 precursor; protein IIIa; protein V; protein pVI; protein pVII; protein pVIII.
SOURCE Human adenovirus type 5
ORGANISM Human adenovirus type 5
REFERENCE 1 (bases 1 to 35935)
AUTHORS Chroboczek J., Bieber F. and Jacrot B.
TITLE The sequence of the genome of adenovirus type 5 and its comparison with the genome of adenovirus type 2
JOURNAL Virology 186 (1), 280-285 (1992)
MEDLINE 92087470
FEATURES Location/Qualifiers

source
CDS
1..35935
/organism="Human adenovirus type 5"
/db_xref="taxon:28285"
CDS
11565..12297
/note="52/55 K protein gene"
/codon_start=2
/protein_id="AAA96406.1"
/db_xref="GI:209843"
/translation="SPNNHRTVLAREVAIGLMHWDVFSALAEQNPNKPLMAQLFL
IVQSRNEAFRALLNIVEPGRWLLDNLINLOSIVQERSLSADKAAINYSMLS
LGKPYAKIYHTFYPIDKVEKIEGYMRMALKVLTLSDDLGYVNRERHKAVSVSR
RELSRELMSLQRLAGTSGDREAESYFDAGADLRWAPSRRALEAAGAGPCLAVAP
ARAGNVGGVEYEDDEYEDCEY"
12318..14075
/codon_start=1
/product="protein IIIa"
/protein_id="AAA96407.1"
/db_xref="GI:289094"
/translation="MMQDATDPAVRAALQSOPSLNSTDDWROVMDRIMSLTARNPDA
FRQPOANRLSAILAVPARANPTHEKVLIAVNAENRAIPDEAGLVYDALLQRY
ARYNSGNVOTNLDLVGDVREAVAQERAAQQQNGLSMVNALFLSTQPANVPERGED
YTNFVSALRLMWTETPQSEYQSGPDFFQTSRQGLQTNLSQAFKNLQGLMGRAPT
GDRATVSSLLTPNSRLLLIAPFTDSSVSRDTYLGHLTLTYREAIGAHVDEHTFO
EITSVSALGQEDTGLEATLNLNRRQKIPSLHSLNSEERILRYVQSVSLNLM
RDGFTPSVALDMTARNMEPCMYASNPFINRLMDYLHRAAVNPVFTTALNPHWLP
PGFYTGGEVPEGNDGFDWDDIDDSVFSPOPTLLELOOQEAALAKESFRPSP
LSDLGAAPRSDASSPFLISGLSTSTRTRPRLLGEESYLNLSLLOPQRNLPFPF
PNNGLISLVDMKSRWKTIAQERHDPVPPRPPTTRQRHQRGLWEDDSDSDSSVLD
LGSGNFFAHLRPLGRMF"
16286..16474
/codon_start=1
/product="protein pVII"
/protein_id="AAA96408.1"
/db_xref="GI:209845"
/translation="MRAARRLAAGIVTVPPRRRRRAAAAAAISAMTQGRGNVYTW
RUSVGLRVPVTRPPRN"
16544..17650
/codon_start=1
/product="protein V"
/protein_id="AAA96409.1"
/db_xref="GI:209846"
/translation="MSKRRKKEEMQVIAPEIYGPKEEQDYKPKRLKRVKKKKDD
DELDDEVELLHATAPRQVQWKRVRVLPPTVTVTPGERSTRITYKRYDEVYD
DELLQANERLGEAYGKRKMDMLALPDEGNPTPSLKPVTLOQVLPALAPSEKRG
LKRESGLAPTQVLMQRORLEDEKMTVEPGLPEVVRVPIKQVAPGLAQVTDV
QIPTTSSATATGEMTQSPVASADAAVQAAVAAAKSTSTEYQDPMFVSA
PRPGRSKYGAASALLPEYALHPSIAPTGGYTYPRRRATRRRTTGTTRRRR
RQPVLAISVRRVAREGRLLVLTPTARYHPSIV"
17774..17920
/codon_start=1
/product="protein 5 precursor"
/protein_id="AAA96410.1"
/db_xref="GI:209847"
/translation="MRAHRRRRASHRRMRGGILPLLIPIIAAIGAVPGIASVALQ
AQRH"
18003..18755
/codon_start=1
/product="protein pVI"
/protein_id="AAA96411.1"
/db_xref="GI:209848"
/translation="MEDINFASLAPRHGRPFMGNWDIGTNSMGGAFSGSLWSGI
KNFSTVNYGSKWNSSTGMRLKLEQFOQKVVVDGLASGIVVDLANQVQNK
INSLDPRPVEEPPAVETVSPGEGKEKRPREDTETLTQIDEPPEALQGLP
TTRTPMATVGLQHTPVTLDLPADTQKPVLPPTAVTVVTRPSRSLRAASGP
RSLRPVAGSNQSTLNSIVGLGVQSLKRRCF"
CDS
25819..26484
/note="100 K protein"
/codon_start=1
/protein_id="AAA96412.1"
/db_xref="GI:209849"
/translation="ESQIIGTFELQGPSPDEKSAAPGLKTPGLMTSAYLRKFPEDY
HAHEIRVEDOSRPPNAELTACVITQGHILGQLQAINKARQEFLLRGRGYLDPQSG
EELNPIPPQYQQPFRALNSQDGTQKRAAAAAATHRGILLGQSGRGGFGGGGGH

[illegible]

ORGANISM	Unknown.
REFERENCE	Unclassified.
AUTHORS	1 (bases 1 to 8299)
TITLE	Wilson,J.M., Fisher,K.J. and Gao,G. Method for improved production of recombinant adeno-associated viruses for gene therapy
JOURNAL	Patent: US 5756283-A 2 26-MAY-1998;
FEATURES	Location/Qualifiers
source	1..8299
BASE COUNT	1816 a 2130 c 2108 g 1963 t 282 others
ORIGIN	/organism="unknown"
Query Match	23.2%; Score 80; DB 5; Length 8299;
Best Local Similarity	100.0%; Pred. No. 2.4e-14;
Matches	80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	266 atgactacgtccggcggttccatttgccatgcatacgaaccacacgatctcggtttct 325
Ddb	3402 ATGACTAGTCGGCGGTCCATTGGCATGCATCTACGACCACGATCTCGGTGCT 3461
QY	326 cggcgcactccgtcacagtag 345
Db	3462 CGCGCACTCCGTCAGTAG 3481
RESULT	5
S82508	LOCUS
DEFINITION	S82508 3119 bp DNA VRL 03-DEC-1996
ACCESSION	9E4:orf2...orf7 [adenovirus type 9, Genomic, 6 genes, 3119 nt].
VERSION	S82508
KEYWORDS	S82508.1 GI:1699390
SOURCE	Human adenovirus type 9.
ORGANISM	Human adenovirus type 9
REFERENCE	Viruses: GDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
AUTHORS	1 (bases 1 to 3119)
TITLE	Jamvier,R. and Shenk,T. Mammary tumors induced by human adenovirus type 9: a role for the viral early region 4 gene
JOURNAL	Breast Cancer Res. Treat. 39 (1), 57-67 (1996)
MEDLINE	96327849
REMARK	Genbank staff at the National Library of Medicine created this entry [NCBI gi 3119390 gb 178541] from the original journal article. This sequence comes from Fig. 1.
COMMENT	Region: 9E4.
FEATURES	Location/Qualifiers
source	1..3119
gene	/organism="Human adenovirus type 9" /db_xref="taxon:10527" 471..848 /gene="E4" /note="early region 4" 471..848 /gene="E4" /note="orf1 transforming protein; This sequence comes from Fig. 1"
CDS	/codon_start=1 /protein_id="AAB37504.1" /db_xref="GI:1699391" /translation="MAESLYAFDTPGGIAPVOEGTSNRYTFPCSEFHPHGVLL HLKVSIVDPVPGYGCRFWALNDYARDILQTSDVIFAGRQELTVLLFNHTDFLVRK 889..1281 /gene="orf2" 889..1281 /gene="orf2" /note="This sequence comes from Fig. 1" /codon_start=1 /protein_id="AAB37505.1" /db_xref="GI:1699392" /translation="MLORRGVSHIVPVGVLVTLEDFTSIOMIKELPRFTTHILEL
gene	889..1281
CDS	889..1281

Qy	241	cgcacctgttgcttcaggaataatgactacgctcg 278
Db	1871	AGGCCCATTTCTGCTGCAGAGTACGACAGGTACG 1908
RESULT	6	
LOCUS	AF108105/c	
DEFINITION	Human adenovirus type 17 complete genome.	15-MAR-1999
ACCESSION	AF108105	
VERSION	AF108105.1	
KEYWORDS	GI:4416335	
SOURCE	Human adenovirus type 17.	
ORGANISM	Human adenovirus type 17	
REFERENCE	1. (bases 1 to 35100)	
AUTHORS	Chillon, M., Bosch, A., Zabner, J., Lav, L., Armentano, D., Welsh, M. J. and Davidson, B. L.	
TITLE	Group D adenoviruses infect primary central nervous system cells more efficiently than those from group C	
JOURNAL	J. Virol. 73 (3), 2537-2540 (1999)	
MEDLINE	99139039	
REFERENCE	2. (bases 1 to 35100)	
AUTHORS	Armentano, D. PhD and Souza, D. W.	
TITLE	Direct Submission	
JOURNAL	Submitted (20-NOV-1998) Molecular Biology, Genzyme Corporation, 31 New York Ave., Framingham, MA 01701-9322, USA	
FEATURES	Location/Qualifiers	
source	1. .35100	
organism="Human adenovirus type 17"		
/db_xref="taxon:46922"		
complement(4991..8266)		
/codon_start=1		
/product="DNA polymerase"		
/protein_id="AAD20327.1"		
/db_xref="GI:4416343"		
/translation="MQCVDTATNSPVEIKYHLKHALTRLYEVNLTLPDLDLRDT MDSQALRYALFRRPRAETWTLPRLGLVSLVLEPOGESHAGENHQGPPLIKF LKGRALVDEVPQVCEYCGRYKHOEGSVRRDPYFHHINSHSSNMWQETQFP PIGSHPRTERLTVDYVENYTWGSGKGLVFMVKPFGDPPELVALARDLAARLW DRWERDPEYCVTPKMAVGQQLFRDDELQTLMAELWASFQANPHLOEWALEQH GLOCPDELTEELKPLHNGRPFMEFLVIGHNINGDEIVLAAQVNNRASYVPGPF RTNRNMPRAKGLFNDFTFALPNSLKRKTDTEFWHGCGDSDSKYQFLKVMVRDT FAULTSLRKAQAYALPERKCCPYKAVNHFMGLSVRADRDGFPFLREYWKDDEYA LNRELWKEAGEAYDIIRETLTYCAMDLVLTAEFAKLQDSYAFHFDRLPAHFN IFORTISSNSHAIFROIIVRAEOQPNFGGLGSPSHLEYEVYRASIRGRCYPTN IGFSPIYVIDICGMVASALHPMPWGPPLNRYEAAAREWQALDASKDIYDF KLCIGLFIIDADPEHLLDVLPPFCRKGRLCWTNEPLRGEVATSDVLTLLHNRG WRVRIYDNTETVFPKWKAREYVOLNIAKERADKDNKTRSAKLNSALYGSF ATRYKLDNKKIVDSQMDSELKSIAAGANIKSSSFLETDNLSAEVMPALREYLPQOL ALVDSAESEDEHPAPETPPSGTQGHVATYKPIFDLAEEDGDCMLTVEKVDPL VNDNRYSHVASFLAWTAFYSEWSEFLYEDRGTPLQDRPIKSVYGDTSFLVBER GHRMLTRCKRIKNGGKLVDPQPELTLVCEYVCAHCGADAFAPSEFLAPKL YALOSLLCPACCKSSGKLRAGHAAELNTELMVNYCLADAQGEDRAREFSTRMSLK RULASQAQGAHFFVTETTLTKLRPKWDMTLAALDAHRLVPSRSRPNRNEEVCWI EMP"		
complement(8311..10161)		
/codon_start=1		
/product="terminal protein"		
/protein_id="AAD20326.1"		
/db_xref="GI:4416342"		
/translation="MDYELPLNTNWRVREPPRASTTAAGITWMSRYLYGHRMLED LAPGAPATQRFELKQPPPHFVGLVQDYLRTCDNDFVDSRAFSRLRYSEVQPLQTV NWSLMANCTYINTGAXHFRFVDYDFODTLTRVQQAALAEVAVADLALVQPLRGVTV RMEDSASADIERLMDHYTKNLSCQCAQMAERLRIQQAGPKDLVLLATRLRLN AYFNTIISNRNSVHRAATYKLCPCQMDLAFERSDPDLDTALSTPOLIRCI VSALSLNPGDPPHYRETMGGVEFLTRPRGRVATENRRRGEMIERFVDRPVRNR RAAPPPPPPEEIEEVWVEEEDAPGFEREVTRATIAELIRLLEDELTVSARNQ PFNFVADYEWAEHLAEGDISENPLRWIMFFVTHIATTLNLFQRLNRYAFAR HVELNLAQVMAKADGDVVYISRWNESGLGAFSQLMRISNDLAATVVERAGRGDLQ EEEIQFMSEIAYQNSQSGVQTEILRQAPVDAEIDSVELSFRTKVPVVTORRQIQ DVNRVVAHASALRAHQDLRPERHADYVLPPLPAGPEPPLPPGAPRHRF"		

J. Virol. 67 (2), 682-693 (1993)
 93124560
 (sites)
 REMARK
 MEDLINE
 4 (bases 4831 to 10470)
 Shu, L.M., Hong, J.S., Wei, Y.F. and Engler, J.A.
 Nucleotide sequence of the genes encoded in early region 2b of
 human adenovirus type 12
 Gene 46 (2-3), 187-195 (1986)
 87106854
 JOURNAL
 MEDLINE
 5 (bases 1 to 3957)
 van Ormondt, H. and Galibert, F.
 Nucleotide sequences of adenovirus DNAs
 Curr. Top. Microbiol. Immunol. 110, 73-142 (1984)
 85002829
 JOURNAL
 MEDLINE
 6 (bases 1 to 3957)
 Shinagawa, M. and Padmanabhan, R.
 Comparative sequence analysis of the inverted terminal repetitions
 from different adenoviruses
 Proc. Natl. Acad. Sci. U.S.A. 77 (7), 3831-3835 (1980)
 81054665
 JOURNAL
 MEDLINE
 7 (bases 1 to 3957)
 Sugisaki, H., Sugimoto, K., Takanami, M., Shiroki, K., Saito, I.,
 Shimojo, H., Sawada, Y., Uemizu, Y., Uesugi, S. and Fujinaga, K.
 Structure and gene organization in the transformed Hind III-G
 fragment of Ad12
 Cell 20 (3), 777-786 (1980)
 81022638
 JOURNAL
 MEDLINE
 8 (bases 1 to 3957)
 Tolun, A., Alestrom, P. and Pettersson, U.
 Sequence of inverted terminal repetitions from different
 adenoviruses: demonstration of conserved sequences and homology
 between SA7 termini and SV40 DNA
 Cell 17 (3), 705-713 (1979)
 80001962
 JOURNAL
 MEDLINE
 9 (bases 1488 to 3861)
 Kimura, T.
 Structure and sequence analysis of the transforming region E1b of
 human adenovirus type 12
 Sapporo Igaku Zasshi 52, 253-267 (1983)
 200966
 JOURNAL
 MEDLINE
 10 (bases 20966 to 22966)
 Kruijer, W., van Schaik, F.M., Speijer, J.G. and Sussenbach, J.S.
 Structure and function of adenovirus DNA binding protein:
 comparison of the amino acid sequences of the Ad5 and Ad12 proteins
 derived from the nucleotide sequence of the corresponding genes
 Virology 128 (1), 140-153 (1983)
 83277521
 JOURNAL
 MEDLINE
 11 (bases 24334 to 24703)
 Engler, J.A. and van Bree, M.P.
 The nucleotide sequence of the gene encoding protein IVa2 in human
 adenovirus type 7
 Gene 19 (1), 71-80 (1982)
 83054637
 JOURNAL
 MEDLINE
 12 (bases 1 to 530)
 Shibata, H., Zheng, J.H., Koikeda, S., Masamune, Y. and Nakanishi, Y.
 Cis- and trans-acting factors for transcription of the adenovirus
 12 E1a gene
 Blochm. Biophys. Acta 1007 (2), 184-191 (1989)
 89150250
 JOURNAL
 MEDLINE
 13 (bases 1 to 34125)
 Juttermann, R., Weyer, U. and Doerfler, W.
 Defect of adenovirus type 12 replication in hamster cells: absence
 of transcription of viral virus-associated and L1 RNAs
 J. Virol. 63 (8), 3535-3540 (1989)
 89311650
 JOURNAL
 MEDLINE
 14 (bases 1 to 34125)
 Kimura, T., Sawada, Y., Shinagawa, M., Shimizu, Y., Shiroki, K.,
 Shimojo, H., Sugisaki, H., Takanami, M., Uemizu, Y. and Fujinaga, K.
 Nucleotide sequence of the transforming early region E1b of
 adenovirus type 12 DNA: structure and gene organization, and
 comparison with those of adenovirus type 5 DNA
 Nucleic Acids Res. 9 (23), 6571-6589 (1981)
 82105565
 JOURNAL
 MEDLINE

REMARK (sites)
 FEATURES
 source
 Location/Qualifiers
 1. .34125
 /organism="Human adenovirus type 12"
 /db_xref="taxon:28282"
 1. .161
 /function="inverted terminal repeat"
 6
 /citation=[5]
 /citation=[6]
 /citation=[7]
 /citation=[8]
 /citation=[12]
 /replace="a"
 8
 /citation=[5]
 /citation=[6]
 /citation=[7]
 /citation=[8]
 /citation=[12]
 /replace="t"
 9. .407
 /note="crossref Ad12 EPD30061 and EPD11197"
 /function="E1a promoter region (-499 to +100)"
 369
 /citation=[5]
 /citation=[6]
 /citation=[7]
 /citation=[8]
 /citation=[12]
 /replace="t"
 503. .1099
 /note="E1A; transcription activation; early protein;
 alternative splicing; crossref SWISS-PROT:E1A_ADE12,
 P03259"
 /codon_start=1
 /protein_id="CAA51877.1"
 /db_xref="GI:313362"
 /db_xref="SWISS-PROT:P03259"
 /translation="MRTMTPLVLSIQADLLEHLVNFNFNEVPSDDLLVPSLYEL
 YLDVSEAGEDNNEQVAEVEFFESLILAASEGFLPEPVLSPVCEPIGSGCMPQLHP
 EDMDLCEMFGPSCDSDEQDENGMAHVSAASAAAADREEREFOLDHPDPGHNCK
 SCEHRTGTNTDMLCCLYLRAVNMFIYSCAMGGGR"
 1029. .1628
 /note="crossref Ad12 EPD07152"
 /function="E1b promoter region (-499 to +100)"
 1542. .2033
 /note="crossref SWISS-PROT:E1BS_ADE12, P04492"
 /codon_start=1
 /product="E1B protein, small T-antigen"
 /protein_id="CAA51878.1"
 /db_xref="GI:313363"
 /db_xref="SWISS-PROT:P04492"
 /translation="MELETVLQSFQSVRQLQYTSKNTSGFWYLFGLSTLSKVVRVK
 EDYRENFENILADCPGLLASLDLCYHLVFEKVVRSIDFSVGTVAASIAFLAILDK
 WSEKSHWDYMLDYNMQLRAWLKRRCVCIYSLARPLTWPLPTLQEEEREENPAV
 VEK"
 1740
 /citation=[9]
 /replace="t"
 1742
 /citation=[9]
 /replace="g"
 1847. .3295
 /note="crossref SWISS-PROT:E1BL_ADE12, P04491"
 /codon_start=1
 /product="E1B protein, large T-antigen"
 /protein_id="CAA51879.1"
 /db_xref="GI:313364"
 /db_xref="SWISS-PROT:P04491"
 /translation="MEREIPELGLHAGLHVNAAVEGMAEEGLHLLAGAAFDHAAA
 DVARGEGGAGPCGGEVNMQVQGHVLDSEGSACDDDKQKESLSLKEAAVLS
 RLTVNLSRPRLETIVYQELQDEFQGMHLYKYSEQLKTHWDFWEDMECAIKAF


```

/translation="MRMLPFFFTGNWDDMFQGLLETEYVDFPEPSEASEMSLHDLF
DVEYDGEEDANQAVDGPISDAEGESGSPEDTFFPHPLTATPHGIVTRIPCRVS
CRRPFAVECIEDLLEEDFTDEPLNLSLRKCS"
<486..1011
/standard_name="E1A 249R"
/number=1
<486..927
/standard_name="E1A 221R"
/number=1
<486..669
/standard_name="E1A 135R"
/number=1
join(486..927,1088..1311)
/standard_name="E1A 221R"
/codon_start=1
/protein_id="AACI3950.1"
/db_xref="GI:303972"
/translation="MRMLPFFFTGNWDDMFQGLLETEYVDFPEPSEASEMSLHDLF
DVEYDGEEDANQAVDGMPEVGERLLSEAEASGSDGSGVEILLPVLDLKLKYEDG
LPSPDPTDPTATAEAEAAPTVYNNENELVLDPENPGRCRACFISDAGESESG
SPEDTFFPHPLTATPHGIVTRIPCRVSRRAVECIEDLLEEDFTDEPLNLSLRK
CS"
<603..603
/notes="E1A-E1B cotranscript"
/number=1
/evidence=experimental
1088..>1311
/standard_name="E1A 221R"
/number=2
1088..>1311
/standard_name="E1A 249R"
/number=2
1088..>1311
/standard_name="E1A 135R"
/number=2
1321..1326
/notes="for E1A"
1373..1378
/notes="for E1B"
1403..>1403
/product="E1B"
/evidence=experimental
1407..>1407
/notes="E1A-E1B cotranscript"
/number=2
/evidence=experimental
1412..>1412
/product="E1B"
/evidence=experimental
<1414..3151
/standard_name="E1B 19K"
/number=1
1414..1917
/standard_name="E1B 19K"
/codon_start=1
/evidence=experimental
/protein_id="AACI3957.1"
/db_xref="GI:303979"
/translation="MELWSELQSQYONLRRLLELASARTSCWRLLFGSTLTNVIYRAK
EYISRFADLLSHNPGFASLNIGHHFFOEVIRNLDSSPGRTVSGLAFCIFILDO
WSAOTHLISQGYTLDYMAWLWRLLRKRVLGCLPAQRPHGLDPVQEEEEENLRAG
LDPSTEL"
<1719..3151
/standard_name="E1B 55K"
/number=1
1719..3149
/standard_name="E1B 55K"
/codon_start=1
/evidence=experimental
/protein_id="AACI3958.1"
/db_xref="GI:303980"
/translation="MERPNSSVAGLYSGLHGNSVENLATEEBSGLRLAGAAARSFGS
SAGRGGGGGPEGRPGIVTPEPDECTSSCGGGINGORGTKRKKNENGEDFLK

```

CDS		ELILSLMSRRHHSVWVALEDFKNGEMNLKYTFEQLKTHLWEAFEDFELALNTFI AKVALRDTYITIKKVNRIRKCAVLGNAVFQTCDRVACAMQSGLPGIIGMSG VTFMNRFVVFGNFVFASSTQTLLIHGVFFONCSGICVDWSGRVSARGCTFVACWKG VGGRNSQMSKKCVFERCIMAAMVEGOARIRHNAGSDNVCPILLKGTTASVKHMTCG GGHSOLLTADGNQALRVHYVSHRPWPVFEHNMARCIVHLGARGMSPSQSN FCHTKYLMEVDASFVRWNNGVEDTLIMELFKVRYDESKVRCPCECGANHINLIYPATL NVTEQLTDHOMMSCLRTDYESSDED join(1719..1938,3061..3149) /standard_name="ElB 15k" /codon_start=1 /evidence=experimental /protein_id="AAC13951.1" /db_xref="GI:303973" /translation="MRPNSSVATGESLGHNGSVENLATBEEGRLLAGAASARFGS SAGRCGGGEPEGRPPFNIGVTEPDPEEATLNVTQEUTDHOHMSCLRTDYESSDED"
exon	<1719..1938	
exon	/number=1 3061..3151 /gene="ElB" /standard_name="ElB 15k" /number=2 3061..3151 /gene="ElB" 3176..3181 /note="for pix" 3207..>3207 /product="pix" /evidence=experimental 3208..>3208 /product="pix" /evidence=experimental 3217..3615 /standard_name="pix" /codon_start=1 /protein_id="AAC13959.1" /db_xref="GI:303981" /translation="MSGFTEGNAVSEFGVFSPLYLTTRLPWSAGVRONYGSNVDDGRP VAPANSTLTIVATIGSGTSVDTAAAAASAATASTARMAADFGLYNQLAASRLREDALS"	
gene		
TATA_signal		
mRNA		
mRNA		
CDS		
Query Match	10.7%; Score 36.8; DB 16; Length 34214;	
Best Local Similarity	44.7%; Pred.No. 0.79;	
Matches 143; Conservative	0; Mismatches 177; Indels 0; Gaps 0;	
QY	1 atggtttcttcagcgtcttccgcctccctcgctgtgtgactcgcgaacgaatgtgttaggt 60 	
Db	33037 ATGCCTCTCCGCTCTCTCCCTCCACCTCCAGTGAGTCGGGATCAGGCCCTTTGTATTCT 32978	
QY	61 tggctgggttggttgtattctcggtggtgtgatgtatcaggcgagcgcgcatgaagga 120 	
Db	32977 TGGTTGGAGTTGGCTTTAACTGCTGTTTGAACGCTTTATGGGGACATAATAAGGTATAAT 32918	
QY	121 gttaccatagaaccgaagcagggggcgctggatgctttgagagagtgaatactac 180 	
Db	32917 GTCAGCATATCCACGGGCTCAGGAGCTGTTAAGTGCATGCGAGTAGTGCTGAACGTT 32858	
QY	181 aactactacagacgcatctaagcggcgagacggagacgacagatctgtttgttcacgcc 240 - - -	
Db	32857 GCCCTGAAACTGAAGAAGCGCGCACACAGAAACCGTCGATTTCGGAACATTTGTTGGCAG 32798	
QY	241 cgcacctggtttgttcaggaaatatgactacgtccggcggttccatttggcgacacct 300 	
Db	32797 CGTGTTCGTTCTTAAAACAGAGTACGAGCTGCGCGAGCTTTTAATTTATGATGCA 32738	
QY	301 acgaccaaacacgatctcgggt 320 	
Db	32737 ACGGACAGATGGTTTCGCT 32718	

RESULT 10
HSL09F14
LOCUS

PRI 23 - NOV - 1999

PRI 23 - NOV - 1999

/product="dJ109F14.1.1 (Transcriptional Enhancer Factor
 TER-5)(isoform 1)"
 /protein_id="CA83625.1"
 /db_xref="GI:4490420"
 /translation="ASNSWASSPCGEAREDEGEGDKGLDNDAGVSPDIEQSQE
 ALAIYPCGGRKILISDEKMGVGRNELIARYIKLRGTRKQVSSHIOVLARKVR
 EYGVGKAMNLDVSKDLQSMASSSACIVASVQNKFPSPPLQAQVSTSSRF
 WSPPLGQGPQSDIKKFAQPAYPIPLPTLSSYEPLAPLPSAAASVPMQDRT
 IASRLRLLEYGSPVQVQDPTYSKHLFHCIGQNPAPFSDPLEADVQVQYDFPE
 KQGLKELKGPNAFFLVKFWADLNSTIQSGPAGVGVSSOYSSADSMITSVSTKV
 CSPGQVKEVETEVARELVREVRIRHSRPMCEYMFHKLKLPKPKYMNVSLEN
 FTIQVTSRDSQETLLVIAFVEVSTSEHGQAHVYKLVKD"
 9406..10064
 /gene="dJ109F14.1"
 /note="predicted CpG island"
 complement(12114..12232)
 /note="MIR repeat: matches 211..90 of consensus"
 complement(13299..13434)
 /note="AluJo repeat: matches 136..1 of consensus"
 13801..14075
 /note="AluX repeat: matches 22..296 of consensus"
 14303..14508
 /note="MIR repeat: matches 28..261 of consensus"
 15438..15559
 /note="FIAM.A repeat: matches 2..127 of consensus"
 16887..17186
 /note="AluSg repeat: matches 1..300 of consensus"
 complement(17298..17368)
 /note="MIR repeat: matches 128..46 of consensus"
 17910..18225
 /note="AluJo repeat: matches 1..301 of consensus"
 18945..19239
 /note="AluSg repeat: matches 2..298 of consensus"
 22414..22439
 /note="13 copies 2 mer aa 92% conserved"
 22927..23234
 /note="AluY repeat: matches 1..298 of consensus"
 23235..23747
 /note="match: GSS B94568"
 complement(23583..23637)
 /note="MIR repeat: matches 88..34 of consensus"
 complement(24131..24272)
 /note="AluX repeat: matches 280..134 of consensus"
 24323..24609
 /note="AluY repeat: matches 2..294 of consensus"
 24773..24922
 /note="AluSg repeat: matches 127..297 of consensus"
 complement(join(25280..25484,25713..25885,26535..26683,
 27037..27117,27191..27265,27625..27644))
 /gene="dJ109F14.2"
 /note="match: cDNAs U12404 U12403 X93352 Z23090; match:
 ESTs AA134024 AA148100 AA531410 AA070054 F17880 AA641205
 AA745013 AA644024 AA720704 AA888465 AA657588 AA146904
 Query Match 10.5%; Score 36.2; DB 10; Length 170245;
 Best Local Similarity 53.1%; Pred. No. 1.4;
 Matches 77; Conservative 0; Mismatches 68; Indels 0; Gaps 0;
 QY 39 ctgcgacgaacgaatgttagtgggtggtggtggtggttattctcggtggtggtat 98
 Db 115386 CTCTACGAAAAATTTAAAAATAGCCGGGTGTGCTGTGCTAGTCCAGCTAC 115445
 QY 99 cagggcagcgcgcgatgattacatagaccgacccgagcggcgccctggtatgc 158
 Db 115446 TCGGAGGCTGAGCAGGAGGAGCCCTGCTGAGCCCGAAGGTCGAGGTCGAGTGAGC 115505
 QY 159 tttagagagtggtatatactacaac 183
 Db 115506 CGAGATTGCTACTGCACTACAGC 115530

RESULT 11
 MNGFAPD/C

LOCUS MNGFAPD 9971 bp DNA ROD 10-FEB-1999
 DEFINITION Mouse gene for glial fibrillary acidic protein (GFAP).
 ACCESSION X02801
 VERSION X02801.1 GI:51065
 KEYWORDS Alu repetitive sequence; direct repeat; glial fibrillary acidic protein; intermediate filament; repetitive sequence; tandem repeat.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 Balcarek, J.M. and Cowan, N.J.
 1 (bases 1 to 9971)
 TITLE Structure of the mouse glial fibrillary acidic protein gene: implications for the evolution of the intermediate filament multigene family
 JOURNAL Nucleic Acids Res. 13 (15), 5527-5543 (1985)
 MEDLINE 85297756
 FEATURES
 source
 1..9971
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 163..164
 /note="put. CAAT-box"
 217..219
 /note="put. TATA-box"
 247..712
 /note="exon 1"
 247
 /note="cap site"
 297..8595
 /gene="GFAP"
 join(297..712,1563..1623,1805..1900,2833..2994,3780..3905,
 4114,4334,5408..5451,7821..7906,8551..8595)
 /gene="GFAP"
 /codon_start=1
 /product="glial fibrillary acidic protein"
 /protein_id="CAA26571.1"
 /db_xref="GI:51066"
 /db_xref="MGD:MGI:95697"
 /translation="MPPRWGSGSGFSQLGTMPFRFSRMTPTLPARVDFSLAGALN
 AGFKTRASERAEEMELNDRFASYIEKVRFLQQNKALAEINQLRAKENPLADVYQ
 AELRELRDLQLTANSARLEVERDNFADLTGTLKQLQDETINLRLEENLAAYVQ
 AHEATLARYDLERKVESLEETQFLKYEVEERDLRQLAQGVHVMQVADYKDLTA
 ALREITQYEAVATSNMOETETWYKSFADLTDAASRAELLRQAKHEANDYRQLQA
 LTCDESLRGTNESLEROMRQEEHAREASAYQALARLEEGSLKEEMARHLQEI
 HDLNVKLAIDIEIATYRKLEGENRTIIPVQTFNSLQIRETSLLDTKSVSEGHKRN
 IVYKIVMRDGEVINDSKQEHKDVVM"
 713..1562
 /gene="GFAP"
 /note="intron I"
 1563..1623
 /gene="GFAP"
 /note="exon 2"
 1624..1804
 /gene="GFAP"
 /note="intron II"
 1805..1900
 /gene="GFAP"
 /note="exon 3"
 1901..2832
 /gene="GFAP"
 /note="intron III"
 2377..2423
 /note="dinucleotide repeats (TG)6 (TC)9 (AC)8"
 2531..2659
 /gene="GFAP"
 /note="Alu sequence"
 2833..2994
 /gene="GFAP"
 /note="exon 4?"
 2995..3779
 /gene="GFAP"
 /note="intron"
 intron
 mRNA
 intron
 mRNA
 intron
 repeat_region
 misc_feature
 mRNA
 intron

REFERENCE AUTHORS TITLE JOURNAL MEDLINE FEATURES	1 (bases 1 to 773)		Location/Qualifiers				
	Yen, P.H., Marsh, B., Allen, E., Tsai, S.P., Ellison, J., Connolly, L., Neiswanger, K. and Shapiro, L.J.						
	The human X-linked steroid sulfatase gene and a Y-encoded pseudogene: evidence for an inversion of the Y chromosome during primate evolution						
	Cell 55 (6), 1123-1135 (1988)						
source	89077541		1..773				
	/organism="Homo sapiens"						
	/db_xref="taxon:9606"						
	/map="Xp22.32"						
intron	<1..265		/gene="STS"				
	/note="G00-120-393"						
	266..689						
	/partial						
CDS	/gene="STS"		/standard_name="STS-X"				
	/codon_start=3						
	/number=5						
	/db_xref="GDB:G00-120-393"						
intron	/product="steroid sulfatase"		/protein_id="AA80598.1"				
	/db_xref="GI:338607"						
	/translation="KWLHGMSCSKTDFCHHPLHGHNFYFGISLTNLRDCKPQEGSV FTTGFKLVFLPQIVGVTLTLAALNCLGLLHVPLGVFVSLLEFLAALILTLFLGLFLH YFRPLNCFMERNYETIIQPMYSYDNLTORLTVEAAQFIQR"						
	690..5773						
BASE COUNT ORIGIN	/gene="STS"		/note="G00-120-393"				
	169 a						
	180 c						
	187 g						
Query Match 10.2%; Score 35.2; DB 9; Length 773;							
Best Local Similarity 58.7%; Pred. No. 1.9; Mismatches 0; Gaps 0;							
Matches 61; Conservative 0; Indels 43; Indels 0; Gaps 0;							
QY	239 ccgcacccgtgtttgtcttcaggaaatgactacgtccggcggttcatttggtgcacgaca 298		Db				
	584 CCTGAACTGCTTCATGATGAGGAACTACGAGATCATTCAGCAGCCCATGCTCATGACA 643						
QY	299 ctacgaccacacgactcgtgtgtctcggcgacacccgtacag 342		Db				
	644 ATCTCACCAGAGCTAACGGTGGGAGCGCGCCAGTTCATACAG 687						
RESULT 13							
HUMSTSB 2401 bp mRNA PRI 13-JAN-1995							
LOCUS Human steroid sulfatase (microsomal), complete cds.							
DEFINITION J04964							
ACCESSION J04964.1 GI:338564							
VERSION glycoprotein; membrane-associated glycoprotein; steroid sulfatase.							
KEYWORDS Human placenta, cDNA to mRNA, clone 5'3.							
SOURCE Homo sapiens							
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.							
REFERENCE 1 (bases 1 to 2401)							
AUTHORS Stein, C., Hille, A., Seidel, J., Rijbnout, S., Waheed, A., Schmidt, B., Geuze, H. and von Figura, K.							
TITLE Cloning and expression of human steroid-sulfatase. Membrane topology, glycosylation, and subcellular distribution in BHK-21 cells							
J. Biol. Chem. 264 (23), 13865-13872 (1989)							
JOURNAL 89340479							
MEDLINE Draft entry and computer-readable sequence for [1] kindly submitted by C.Stein, 01-JUN-1989.							
COMMENT Location/Qualifiers							
FEATURES 1..2401							
source /organism="Homo sapiens"							
/note="intron IV"							
3780..3905							
/gene="GFAP"							
/note="exon 5"							
3906..4113							
/gene="GFAP"							
/note="intron V"							
4114..4334							
/gene="GFAP"							
/note="exon 6"							
4335..5407							
/gene="GFAP"							
/note="intron VI"							
5142..5208							
/note="homopurine sequence (GA)20 (GAA)9"							
5217..5272							
/note="GT repeats (GT)24"							
5408..5451							
/gene="GFAP"							
/note="exon 7"							
5452..7820							
/gene="GFAP"							
/note="intron VII"							
5761..5822							
/note="TC repeats (TC)25"							
5948..6143							
/note="region with 7 imp. direct repeats (28 bp)"							
6260..6943							
/note="region with 19 imp. direct repeats (34 bp)"							
7907..8550							
/gene="GFAP"							
/note="intron VIII"							
8184..8221							
/note="(AT)13 sequence"							
8551..8551							
/gene="GFAP"							
/note="exon 9"							
9925..9929							
/note="put. polyadenylation signal"							
BASE COUNT 2480 a 2427 c 2672 g 2392 t							
ORIGIN							
Query Match 10.4%; Score 36; DB 12; Length 9971;							
Best Local Similarity 55.6%; Pred. No. 1.3; Mismatches 0; Gaps 0;							
Matches 69; Conservative 0; Indels 55; Indels 0; Gaps 0;							
QY	123 ttacatagaacccgaagcagggcgctggtgatctttgagagagtggtgatactacaa 182		Db				
	6261 TTACACAGAGCCAGGACCTAGGCTGGGCTGTGAGCTTCTAGACAAAGGACATGGTTCAG 6202						
QY	183 ctactacacagagcagctatcagcgcgagacgagacgagatctgttcacgccc 242		Db				
	6201 TCCCTAGAGGAGCGCTGGGAGGAAGACAAAGCTGGNAAGGTGTGCTGNAATGCCG 6142						
QY	243 cacc 246		Db				
	6141 CGCC 6138						
RESULT 12							
HUMSTXG1 773 bp DNA PRI 13-JAN-1995							
LOCUS Human chromosome X steroid sulfatase (STS-X) gene, exon 5 and flanking regions.							
DEFINITION M23945							
ACCESSION M23945.1 GI:338604							
VERSION steroid sulfatase.							
KEYWORDS 1 of 2							
SEGMENT Human blood, cultured fibroblasts or tissue DNA.							
SOURCE Homo sapiens							
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.							

AUTHORS	Yen, P. H.
JOURNAL	Unpublished (1988)
COMMENT	There is a steroid sulfatase (STS) pseudogene on Yq.
FEATURES	Location/Qualifiers
source	1..6520
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/map="Xp22.32"
mrna	<1..2429
	/note="steroid sulfatase mRNA"
sig_peptide	221..285
	/gene="STS"
gene	/note="steroid sulfatase signal peptide"
	221..1972
	/gene="STS"
CDS	221..1972
	/gene="STS"
	/note="steroid sulfatase precursor (EC 3.1.6.2)"
	/codon_start=1
	/db_xref="GDB:G00-120-393"
	/protein_id="AAA60596.1"
	/db_xref="GI:338514"
	/translation="MPLRKMKRIPPLLLFELWEAESHEASPNLLIVLVA YNGKTRTPNIDRLASGGVKITQHLAASPLCTPSRAFMTRGTPVRSRG LFTSSGGLPDEITFAKLLKDCGYSTALICKWHLGMSCHSTDFCHH GISUNLNDCKPFGSGVFTGPKRNVLPQIVGVTLTLTAAALNCGLD LLFLAALILFLGLHYFRPLNCFMNRNTEIIOQPMYSNDLNTQRTVLT EFPFLVLSYLHVHTALFSSKDFAGKSGHYGDAVEEMDWSVQGLIN TLIFTSQQGAHVVESSKGEIHGSGNGIYKGGKANNWEGGIRPGILL KIDETSNMDFLPTFAKLAGAPEDRIIDGRDMLPLLEKGSORSHPHE NAVWRHQPNSITSINKAFCTFPNTNPVSGNCEATHVCFGCGSVYTHHDD PRENPILTAPSEPRFVILLKVMQEAADRHQTILFEPVDFGSNNFLWKK GLSCQCQREKQDKLSR" ;
mat_peptide	287..1969
	/gene="STS"
	/note="steroid sulfatase"
BASE COUNT	1881 a 1440 c 1363 g 1835 t
ORIGIN	Chromosome Xp22.32. 1 others

Query Match	10.2%;	Score 35.2;	DB 9;	Length 6520;	
Best Local Similarity	58.7%;	Pred. No. 2.2;			
Matches	61;	Conservative	0;	Mismatches	43;
			Indels		Gaps
					0;
QY 239	ccgcacccgtgtttgcttcaggaatatgactacgtccgcgcttcattggcatgaca	298			
Db 936	CCTGAACTCCTTCATGATGAGGAACATACGAGATCATTCAGACGCCCATGCTATGACA	995			
QY 299	ctacgaccaacacagatctcgggtgtctcggcgcactccgctacag	342			
Db 996	ATCTACCCAGAGGCTAACGGTGGAGCGGCCCCAGTTCATACAG	1039			
RESULT 15					
AC005704/c					
LOCUS	AC005704	98033 bp	DNA	PRI	26-SEP-1998
DEFINITION	Homo sapiens Xp22 bins 29-30 BAC GSHB-227L7 (Genome Systems Human BAC Library) complete sequence.				
ACCESSION	AC005704				
VERSION					
KEYWORDS	AC005704.1	GI:3659406			
HTG					
SOURCE	human.				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 98033)				
AUTHORS	Muzny,D., Aronson,A.D., Adams,C., Brundage,E., Bunac,C., Carvelli,K., Chacko,J., Chen,J., Di,W., Ding,Y., Dugan,S., Durbin,J., Forcum,J., Ganesh,R., Garcia,C., Goodman,M., Gorrell,J.H., Haywood,M., Hernandez,J., Jackson,L., Jin,S., Kampal,K., Karpathy,S., Kovar,C., Leal,B., Li,Y., Lichtarge,O., Liu,W., Logan,O., Lu,J., Ly,T., Martinez,C., Oswal,G., Perez,L., Rashid,N.D., Rowland,K., Savage,L., Scherer,S.E., Shen,H.,				

Simon, M., Stovall, K., Timms, K.M., Todd, J., Vo, O., Williamson, A.,
Worley, K.C., Yu, W., Chinnault, C., Nelson, D. and Gibbs, R.A.

Direct Submission

Unpublished

2 (bases 1 to 99033)

Worley, K.C.

Direct Submission

Submitted (26-SEP-1998) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

Sequencing is completed to a minimum standard of double strand
coverage with a minimum of 2 clones and 2 reads with no ambiguities
or 2 chemistries with a minimum of 2 clones and 3 reads with no
ambiguities. If the sequence quality does not meet this standard,
it will be indicated in the annotation.

The repeat regions shown were identified using RepeatMasker by
Adrian Smit.

Sequence similarities were identified using Powerblast by Jinghui
Zhang.

Exon/Intron boundaries of identified genes were chosen if there
were canonical splice junctions that maintained sequence continuity
across the splice junctions.

Location/Qualifiers

source

1. 99033

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="X"

/map="Xp22"

/clone="GSHB-227L7"

/clone.lib="Genome Systems Human Bac Library"

1065. 1348

/rpt_family="MER7B"

1352. 1649

/rpt_family="AluSx"

1673. 1754

/rpt_family="MER7B"

4011. 4317

/rpt_family="AluSx"

4546. 4841

/rpt_family="Alu"

complement(join(4947. 5371,7132. 7254,7357. 7478,

11384. 11524,45034. 45264))

/gene="Human steroid sulfatase (STS)"

/note="M16505 incomplete sequence."

4990. 5097

/standard_name="DXS6767"

/db_xref="GDB:345311"

complement(6292. 6384)

/rpt_family="(TAGA)n"

complement(6394. 6514)

/rpt_family="(TAGA)n"

complement(6535. 6659)

/rpt_family="(TAGA)n"

complement(6676. 6795)

/rpt_family="(TAGA)n"

6867. 6926

/rpt_family="(CA)n"

complement(7521. 7749)

/rpt_family="AluSx"

9839. 10140

/rpt_family="AluSg"

10550. 10910

/rpt_family="MLT1A1"

complement(11368. 11849)

/rpt_family="AluJo"

complement(12735. 13019)

/rpt_family="AluJo"

complement(13060. 13141)

/rpt_family="L1PB3"

13951. 14026

/rpt_family="L2"

complement(14285. 14315)

/rpt_family="(GAAA)n"

14495. 14773

/rpt_family="AluJo"

15815. 15835

/rpt_family="AT-rich"

15849. 16022

/rpt_family="L1MB3"

complement(16848. 17024)

/rpt_family="MER20"

17654. 17684

/rpt_family="(CA)n"

18115. 18401

/rpt_family="AluY"

18402. 18457

/rpt_family="AluY"

complement(19761. 19828)

/rpt_family="MIR"

complement(19909. 20144)

/rpt_family="MER20"

complement(20635. 20938)

/rpt_family="AluSx"

21356. 21637

/rpt_family="AluJo"

21963. 22072

/rpt_family="MIR"

complement(22109. 22307)

/rpt_family="AluJb"

22497. 23054

/rpt_family="MLT1F"

complement(23968. 24092)

/rpt_family="L2"

complement(24167. 24253)

/rpt_family="(TA)n"

24549. 24838

/rpt_family="AluSg"

25704. 25741

/rpt_family="AT-rich"

complement(25883. 26172)

/rpt_family="AluJo"

26439. 26494

/rpt_family="L1PAL2"

complement(26495. 26541)

/rpt_family="(TAGA)n"

26559. 26587

/rpt_family="(CA)n"

26589. 26674

/rpt_family="L1PAL2"

26755. 27035

/rpt_family="L1PB3"

complement(27042. 27156)

/rpt_family="MLT2D"

27271. 27407

/rpt_family="AluSg/x"

27408. 27518

/rpt_family="MER4_internal"

27530. 28319

/rpt_family="MER4_internal"

complement(28320. 28563)

/rpt_family="L1TR8"

28564. 28849

/rpt_family="MER4_internal"

complement(28946. 29228)

/rpt_family="AluJb"

31178. 31326

/rpt_family="L1M1"

31363. 31628

/rpt_family="AluSx"

complement(31631. 31987)

/rpt_family="THE1C"

31988. 32026

/rpt_family="L1M1"

Search completed: June 23, 2000, 09:36:27
Job time: 12569 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2000, 09:07:32 ; Search time 20.4 seconds
(without alignments)
341.359 Million cell updates/sec

Title: US-09-214-478-2
Perfect score: 1594
Sequence: 1 WTSGVPGMTLRPTRSRLS.....RALLQHRPILMHYDSTPM 294

Scoring table:
Gapop 10.0 , Gapext 0.5

Searched: 188963 seqs, 23686106 residues

Total number of hits satisfying chosen parameters: 188963

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : A_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1589	99.7	534	1 W03197	Glucocorticoid rec
2	1583	99.3	294	1 W01729	Human Ad5 E4 regio
3	101.5	6.4	383	1 W41592	Rat FRAG1 protein.
4	88.5	5.6	191	1 R37813	BamHI restriction
5	75.5	4.7	458	1 W48269	Open reading frame
6	74	4.6	1255	1 W01111	HER-2/neu protein.
7	74	4.6	1255	1 W32406	Human HER-2/neu on
8	73.5	4.6	265	1 W03549	ORF-5 protein sequ
9	73	4.6	391	1 W37324	Cytosolic glycerol
10	73	4.6	391	1 W60255	Klebsiella pneumon
11	73	4.6	391	1 W30680	Glycerol-3-phospha
12	72	4.5	624	1 R08222	Extracellular port
13	71	4.5	369	1 W62647	Mature durum wheat
14	71	4.5	664	1 P02229	Alcohol-oxidase N
15	70.5	4.4	1058	1 R34843	HER4 with alternat
16	70.5	4.4	1058	1 R31734	Receptor tyrosine
17	70.5	4.4	1308	1 R54841	HER4. New recombin
18	70.5	4.4	1308	1 R31733	Receptor tyrosine
19	70	4.4	271	1 W82002	Human foetal brain
20	70	4.4	348	1 R38483	Rhodopsin protein.
21	70	4.4	348	1 R31116	Rhodopsin. Diagnos
22	70	4.4	664	1 R11606	Recombinant alchoh
23	70	4.4	782	1 W19764	Her2-QM-CSF immuno
24	70	4.4	2998	1 W19691	ATM mutant 3403del
25	70	4.4	3001	1 W19692	ATM mutant 4612del
26	70	4.4	3046	1 W19690	ATM mutant G9140T
27	70	4.4	3053	1 W19694	ATM mutant 7836del
28	70	4.4	3054	1 W19695	ATM mutant 7279del
29	70	4.4	3055	1 W19697	ATM mutant 5435del
30	70	4.4	3055	1 W19693	ATM mutant 8578del
31	70	4.4	3056	1 W19698	ATM mutant E2904G.
32	70	4.4	3056	1 W19699	ATM mutant Leu43pr
33	70	4.4	3056	1 W19733	Cell cycle checkpo
34	70	4.4	3056	1 W37133	Ataxia-telangiecta

35 70 4.4 3056 1 W84269 Human ataxia telan
36 70 4.4 3059 1 W19696 ATM mutant 5319ins
37 70 4.4 3066 1 W36178 Murine Ataxia-tela
38 69.5 4.4 266 1 W06833 Polypeptide encode
39 69.5 4.4 662 1 R33304 Canine distemper v
40 69.5 4.4 855 1 R28033 Env polypeptide of
41 69.5 4.4 2963 1 W56444 Fragment HgJ1789 o
42 69 4.3 350 1 R68811 Interleukin-8 rece
43 69 4.3 350 1 R80756 Recombinant high a
44 69 4.3 350 1 R80951 ATM mutant C1240T.
45 69 4.3 413 1 W19682

ALIGNMENTS

RESULT 1
W03197 W03197 standard; Protein: 534 AA.
AC W03197;
DT 16-APR-1997 (first entry)
DE Glucocorticoid receptor/adenoviral ORF6 chimaeric fusion protein.
KW Human adenovirus; inverted tandem repeat; E4; E1: open reading frame;
KW expression; late gene; cell line; defective recombinant adenovirus;
KW complementation; gene therapy; cystic fibrosis; dystrophy; PCR; infection;
KW neurodegenerative disease; blood coagulation disorder; viral infection;
KW primer; amplification; polymerase chain reaction; fusion protein.
OS Homo sapiens/Adenovirus Ad5.
PN W09622378-Al.
PD 23-JUL-1996.
PF 19-JAN-1996; F00088.
PR 20-JAN-1995; FR-000747.
PR 01-JUN-1995; FR-006532.
PR 08-SEP-1995; FR-010541.
PA (RHON) RHONE POULENC RORER SA.
PI Dedieu J, Latta M, Orsini C, Perricaudet M, Vigne E;
PI Yen P, Dedieu JF;
DR WPI; 96-354535/35.
DR N-PSDB; T31320.
PT Cells for prodn. of recombinant adeno and adeno-associated virus -
PT contain part of the E4 region of adeno virus, used to produce new
PT adeno virus defective in E1 and E4 regions for use as gene therapy
PT vectors
PS Example 1; Page 41-43; 69pp; French.
CC The invention relates to novel cell lines which can support the
CC replication and production of defective recombinant adenoviruses which
CC lack all or part of the E1 and E4 regions. The novel cell lines contain
CC the adenoviral E4 region, esp. the open reading frame 6 and/or 7,
CC inserted into its gene under control of a promoter, pref. from MTM which
CC is regulated by dexamethasone. The E4 ORF6 may also be inserted into the
CC genome as a chimaeric fusion gene comprising a glucocorticoid receptor
CC hormone binding domain fused N-terminally to the adenoviral ORF6 gene.
CC The fusion protein generated (this sequence) is retained in the cytoplasm
CC of the cell in the absence of hormone by interaction of the hormone
CC binding domain with the hsp90 protein or other factors. Contacting the
CC fusion protein with the hormone causes translocation of the protein to
CC the cell's nucleus where the ORF6 product functions. The gene was
CC generated by amplifying the sequence encoding the hormone binding domain
CC with primers T31316-7 and recombining the fragment with the adenoviral
CC ORF6 sequence. The defective viruses generated in the new cells can be
CC used for gene therapy e.g. for cystic fibrosis, dystrophy,
CC neurodegenerative diseases, blood coagulation disorders, viral
CC infections, etc.
SQ Sequence 534 AA;

Query Match 99.7%; Score 1589; DB 1; Length 534;
Best Local Similarity 100.0%; Pred. No. 2.6e-164;
Matches 293; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 2 TTSGVPGMTLRPTRSRLSRRTPYSRDLPPFETETRAILEDHPHLLPECTLMHNSY 61
Db 242 TTSGVPGMTLRPTRSRLSRRTPYSRDLPPFETETRAILEDHPHLLPECTLMHNSY 301

QY 62 YVRLPCSVGFTLLIQEVPVWDMVLTREELVILRKCMHVCLCCANIDIMTSMIHYESWA 121
 DB 302 YVRLPCSVGFTLLIQEVPVWDMVLTREELVILRKCMHVCLCCANIDIMTSMIHYESWA 361
 QY 122 LCHCSPSGLOIAGGQVLAWSFRVVDGAMFNQRIWYREVNNMPKVFMSVFM 181
 DB 362 LCHCSPSGLOIAGGQVLAWSFRVVDGAMFNQRIWYREVNNMPKVFMSVFM 421
 QY 182 RGRHLIYLRWYDGHVGSVVPAMSGYSAHCGILNNIVVLCSSYCADLSEIRVCCARR 241
 DB 422 RGRHLIYLRWYDGHVGSVVPAMSGYSAHCGILNNIVVLCSSYCADLSEIRVCCARR 481
 QY 242 TRRLMLRAVRIIAETAMLYSCRTERRRQOFIRALLQHRPILMHYDSTPM 294
 DB 482 TRRLMLRAVRIIAETAMLYSCRTERRRQOFIRALLQHRPILMHYDSTPM 534
 RESULT 2
 ID W01729 standard; Protein; 294 AA.
 AC W01729;
 DT 27-AUG-1997 (first entry)
 DE Human Ad5 E4 region ORF6 encoded by plasmid pMWVE4ORF6 minigene.
 KW Minigene; human adenovirus type 5; open reading frame; promoter; in vivo;
 mouse mammary tumour virus; growth hormone; terminator; recombinant;
 packaging cell line; adeno-associated virus; ex vivo; gene therapy;
 inherited disease; cancer; genetic dysfunction; cystic fibrosis.
 OS Human adenovirus type 5.
 FH Key Location/Qualifiers
 FT misc_difference 115 /note= "encoded by codon CAT"
 FT misc_difference 191 /note= "encoded by codon TTG"
 FT W09639530-A2.
 PD 12-DEC-1996.
 PF 04-JUN-1996; U10245.
 PR 05-JUN-1995; US-462014.
 PR 27-OCT-1995; US-549489.
 PA (TYPE-) UNIV PENNSYLVANIA.
 PI Fisher KJ, Gao G, Wilson JM;
 DR WPI; 97-043152/04.
 DR N-PSDB; T59270.
 PT Recombinant adeno-associated virus contg. second gene which
 facilitates its conversion from single stranded to double stranded
 virus - enhances efficiency of ex vivo transduction into target cell
 PS Example 1; Page 72-73; 131pp; English.
 CC This is the amino acid sequence of the product of the human adenovirus
 type 5 (Ad5) E4 region open reading frame (ORF) 6. The coding sequence
 was placed under control of the mouse mammary tumour virus promoter and
 a growth hormone terminator sequence and used to create the minigene in
 plasmid pMWVE4ORF6. The minigene is used to generate a packaging cell
 line expressing the E4 ORF6 product which is used to rescue and package
 an E4 deficient recombinant adeno-associated virus (rAAV). The rAAV
 comprises at least 2 selected genes, the first gene able to express the
 gene of interest in the target cell; and the second gene able to convert
 the single stranded virus to its double stranded form upon expression.
 CC The rAAV is useful in pharmaceutical compositions for use in ex vivo and
 in vivo gene therapy for the treatment of inherited diseases, cancer and
 other genetic dysfunctions, e.g. cystic fibrosis.
 SQ Sequence 294 AA;

Query Match 99.3%; Score 1583; DB 1; Length 294;
 Best Local Similarity 99.3%; Pred. No. 5.1e-164;
 Matches 292; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MTGSGVFGMTLRTSRSLRRTPYSRDLRPPETETRAITLEDHPLPCNTLTHNVS 60
 DB 1 MTGSGVFGMTLRTSRSLRRTPYSRDLRPPETETRAITLEDHPLPCNTLTHNVS 60
 QY 61 YVRLPCSVGFTLLIQEVPVWDMVLTREELVILRKCMHVCLCCANIDIMTSMIHYESW 120

DB 61 YVRLPCSVGFTLLIQEVPVWDMVLTREELVILRKCMHVCLCCANIDIMTSMIHYESW 120
 QY 121 ALHCHCSPSGLOIAGGQVLAWSFRVVDGAMFNQRIWYREVNNMPKVFMSVFM 180
 DB 121 ALHCHCSPSGLOIAGGQVLAWSFRVVDGAMFNQRIWYREVNNMPKVFMSVFM 180
 QY 181 MRGRHLIYLRWYDGHVGSVVPAMSGYSAHCGILNNIVVLCSSYCADLSEIRVCCARR 240
 DB 181 MRGRHLIYLRWYDGHVGSVVPAMSGYSAHCGILNNIVVLCSSYCADLSEIRVCCARR 240
 QY 241 TRRLMLRAVRIIAETAMLYSCRTERRRQOFIRALLQHRPILMHYDSTPM 294
 DB 241 TRRLMLRAVRIIAETAMLYSCRTERRRQOFIRALLQHRPILMHYDSTPM 294
 RESULT 3
 ID W41592 standard; Protein; 383 AA.
 AC W41592;
 DT 18-JUN-1998 (first entry)
 DE Rat FRAG1 protein.
 KW Fibroblast growth factor receptor activating gene 1; FRAG1; rat; FGFR2;
 fibroblast growth factor receptor 2; transforming activity; osteosarcoma;
 autophosphorylation activity; chromosomal abnormality detection;
 neoplasia.
 OS Rattus sp.
 FH Key Location/Qualifiers
 FT Peptide 1..129 /note= "signal peptide"
 FT Protein 130..383 /note= "mature FRAG1"
 FT Misc_difference 185 /note= "encoded by GTG"
 FT Misc_difference 245 /note= "encoded by CAC"
 FT W09748813-A2.
 PD 24-DEC-1997.
 PF 18-JUN-1997; U10660.
 PR 18-JUN-1996; US-020009.
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PI Lorenzi MV, Miki T;
 DR WPI; 98-063151/06.
 DR N-PSDB; V04269.
 PT Rat and human FRAG1 nucleic acid and related proteins - used for
 diagnosis of neoplasia, particularly osteosarcoma
 PS Claim 13; Fig 4; 35pp; English.
 CC This sequence is the rat fibroblast growth factor receptor activating
 gene 1 (FRAG1) protein of the invention. When expressed in frame
 with FGFR2 (fibroblast growth factor receptor 2), the FRAG1 gene
 stimulates the transforming (autophosphorylation) activity of FGFR2. The
 FRAG1 gene can be used as a hybridisation probe to detect chromosomal
 abnormalities, particularly a rearrangement that fuses FRAG1 to another
 gene, resulting in expression of a fusion protein. Detection of such
 rearrangements is used to diagnose neoplasia, specifically osteosarcoma.
 CC These probes (or related primers) can also be used to identify related
 genes in other species. Agents that inhibit FRAG1 (e.g. antibodies or
 antisense molecules) may be useful therapeutically and the FRAG1 protein
 can also be used to screen for drugs. Antibodies are used for
 purification of the FRAG1 protein; in cloning FRAG1 alleles and
 homologues from expression libraries, and as probes in protein blots and
 immunoassays.
 SQ Sequence 383 AA;

Query Match 6.4%; Score 101.5; DB 1; Length 383;
 Best Local Similarity 22.2%; Pred. No. 0.0057;
 Matches 66; Conservative 36; Mismatches 106; Indels 89; Gaps 17;
 QY 11 TLRPTRSLRRTPYSRDLRPPETETRAITLEDHPLPCNTLTHNVS 56
 DB 40 SLRPTPERPRLTPLLRLPAPSPGWAGAPGAPLATGVPPP--SHYLGPSAQDPSR 97
 QY 57 HNVSIVRG-----LPC---SVGTLLIQEVPVWDMVLTRE-ELVILRKCM--HV 99

Db 98 H-CGWRGGALEPQTSFSLPCQSLACFESDKMYQVP--LTLDKRGTVLRLFTMVALI 154
QY 100 CLCCANID-----IMTSMIHGYSWALHCHSSPGSLQCIA---GGQVLASWFRWVDGA 152
Db 155 TVCCPLVAFFCIWLSLLFHFKETSTHCGC--PNYLPSSVATGGEV----- 201
QY 153 MNQRFIYREVYNNMPKEYMFMSSVPMRGRHLIYLRWDYDGHVGSVVPAMSGYSAL- 212
Db 201 --PQRYVW-----RFGICGLHSAPELTAFAFW-NHYLSCASPCP--GYRLLC 242
QY 212 -----HCGTLNIVVLCSSYCADLSEIRVRCARRTRMLRAVRIIAETAMLYSC 264
Db 243 RLNFSLNVVENLALLVLYVSSSEDTIHENA-----FVFIASLSYMLLTC 290

RESULT 4

R37813 ID R37813 standard; Protein; 191 AA.
AC R37813;
DT 13-OCT-1993 (first entry)
DE BamHI restriction fragment 1 (BHRFI) gene product.
KW Antisense oligonucleotide; hybridisation; EBV;
KW Infection; infectious mononucleosis; Hodgkin's disease;
KW Burkitt's lymphoma; undifferentiated nasopharyngeal carcinoma;
KW AIDS related lymphoma; angioimmunoblastic lymphadenopathy;
KW EBV-induced proliferative disease; bone marrow.
OS Epstein-Barr virus.
PN W09311267-A.
PD 10-JUN-1993.
PF 04-DEC-1992; G02256.
PR 05-DEC-1991; GB-025891.
PA (CANC-) CANCER RES INST.
PI Cunningham D, Hickish T;
DR WPI: 93-197083/24.
DR N-PSDB: Q39619.
PT New anti-sense oligo-nucleotide which hybridises with BHRFI gene of EBV - for treating Epstein-Barr virus infection, infectious mononucleosis, Hodgkin's disease, Burkitt's lymphoma etc.
PS Disclosure; Page 11-12; 20pp; English.
CC The oligonucleotides given in Q39617-18 are selectively hybridisable to the BHRFI gene (Q39619). The antisense oligonucleotides can be used to treat EBV infection in patients suffering from infectious mononucleosis, Hodgkin's disease, Burkitt's lymphoma, undifferentiated nasopharyngeal carcinoma, AIDS related lymphoma or angioimmunoblastic lymphadenopathy. They can also be used to treat EBV-induced proliferative diseases by treating bone marrow from a patient to destroy proliferating tumour cells, then reimplanting the marrow into the patient.
SQ Sequence 191 AA;

Query Match 5.68; Score 88.5; DB 1; Length 191;
Best Local Similarity 23.38; Pred. No. 0.038;
Matches 49; Conservative 18; Mismatches 66; Indels 77; Gaps 14;
QY 11 TLRTSRSLRTPYSRDLPPFTETRAILEDHPLPEC-----NLTMHN 58
Db 24 TLHPVLELAARETL---RLSPEDI---VLYRVHVEEIIERSEFTETWNRFTHT 76
QY 59 VS-----YVRGLPCSVGTLLIOEWWVPMDVLTREELVILRKCMHVC--LCCA 104
Db 77 ERVLDLNSVLEIFHRGDP-SLGRAL--AWMA-W-----CMHACRTLCN 118
QY 105 N-----IDMTSMI---HGYESWALHCHSSPGSLQCIAGGQVLASWFRWVDGAMFN 155
Db 119 OSTPYVVDLSVRGMLSEGLDGH-IRHQ-----WG-----WSTLIENIPGS 161
QY 156 QRFIYREVYNNMPKEYMFMSSVPMRGRH 185
Db 162 RRFSTWTLFAGLTSLVICSYLFISGRH 191

RESULT 5

W48269 ID W48269 standard; Protein; 458 AA.
AC W48269;
DT 23-JUN-1998 (first entry)
DE Open reading frame RHKI protein SEQ ID NO:6 from J10075789 Example 2.
KW Killer protein; protease resistant; HM-1; RHKI; Hansenula mrakii;
KW yeast; mutant; antibacterial; soap.
OS Unidentified.
PN J10075789-A.
PD 24-MAR-1998.
PF 04-SEP-1996; 234393.
PR 04-SEP-1996; JP-234393.
PA (EIJT-) EIJIN KENKYUSHO KK.
DR WPI: 98-244362/22.
DR N-PSDB: V20655.
PT Hansenula mrakii protease resistant killer protein - has antibacterial properties and may be used as active ingredient in e.g. soap
PS Example 2; Page 10-12; 16pp; Japanese.
CC The present sequence represents RHKI used from an example of the present invention which describes a Hansenula mrakii protease resistant killer protein. The killer protein has antibacterial activity and so can be used in an antibacterial agent which contains the killer protein as the active component. The killer protein can also be used as the active component in an antibacterial soap. The protein, which was derived from the yeast Hansenula mrakii, has antibacterial properties and may be used in sterilising and cleaning compositions e.g. soap.
CC The protein has a high protease resistance and can be prepared recombinantly in a large amount.
SQ Sequence 458 AA;

Query Match 4.78; Score 75.5; DB 1; Length 458;
Best Local Similarity 19.58; Pred. No. 4.9;
Matches 31; Conservative 31; Mismatches 60; Indels 37; Gaps 9;

QY 59 VSYVRGLP-----CSVGQF--TLIQEWVPMVMDVLTREELVILRKCMHVCCLCANIDIMTS 111
Db 261 VPFLRSFPQQLHCAFNFGKFMQWSINQWM---DEAEFNDKRFHLALISHLIALT 317
QY 112 MMINGY-----ESWALHCHSSPGSLQCIAGQ-----VLASWFRWVDGAMFNORF 158
Db 318 LFVTRYPRILPDALSSSLCH---PLRKNVLANPAKTIFFVLIASNFTIGVLSRSLHYQF 374
QY 159 I-WYREVYNNMPKEYMFMSSVPMRGRHLIYLRWDYDGH 196
Db 375 LSWY-----HWTLPILIFWGMPPFVG-----PIWYVLH 403

RESULT 6

W01111 ID W01111 standard; Protein; 1255 AA.
AC W01111;
DT 01-JAN-1997 (first entry)
DE HER-2/neu protein.
KW HER-2/neu; c-erbB1; p185; oncogene; tyrosine protein kinase;
KW breast cancer; ovary cancer; colon cancer; lung cancer;
KW prostate cancer; immunisation; tumour; vaccine; vector.
OS Homo sapiens.
PN domain
FT Key
FT Location/Qualifiers
FT 676..1255
FT /label= Intracellular_domain
FT /note= "claimed domain, useful for immunisation"
ET W09630514-A1.
PN 03-OCT-1996.
PD 28-MAR-1996; U01689.
PR 31-MAR-1995; US-414417.
PA (UNIW) UNIV WASHINGTON.
PI Cheever MA, Disis ML;
DR WPI: 96-455361/45.
DR N-PSDB: T40739.
PT DNA encoding HER-2-neu poly:peptide(s) - used for prevention or

PT treatment of malignancies with which the HER-2/neu oncogene is associated

PS Claim 2; Page 56-61; 71pp; English.

CC Human HER-2/neu protein (W0111), also called p185 or c-erbB2, is the product of the HER-2/neu oncogene (see also T40739). The protein is over-expressed in various cancers, including breast, ovarian, colon, lung and prostate. The intracellular domain of the protein can be used to immunise an animal against a malignancy with which the oncogene is associated. The polypeptide can be produced in transformed host cells for use in immunisation. Alternatively, CC animal cells are transfected in vivo or ex vivo with a viral vector that directs expression of the polypeptide.

CC Sequence 1255 AA;

Query Match 4.6%; Score 74; DB 1; Length 1255;
Best Local Similarity 18.2%; Pred. No. 28;
Matches 64; Conservative 34; Mismatches 117; Indels 136; Gaps 15;

QY 20 SRTPTYSRDLPPPTETTRAT-----ILEDHPLLPCEVTL-----TMHNVSY---VR 63
DB 387 SNTAPLOPEQLQVFTLEETITGLYISAWPDSLPLSVFQNLQVIRGRILHNGAYSLTIQ 446
QY 64 GLPCS-----VGFLLQIE-----WVVPDMLTRELVLIR-----KCMH 98
DB 447 GUGISWLGRLSRLGSLGALIHNNHLCFVHTVPDQLFRNPQALLHTANRDECEVG 506
QY 99 VCLCCANIDIMTSMIHGYESWALCHSCSPGSLQCIAGGQVLASWFRMVVDGAMFNORF 158
DB 507 EGLACHQL-----CARGHCWGPGPTQCVNCSQ-----F 534
QY 159 IWYREVNV-----YMPKEVFMSSVFMGRHLIYLRLWDYGHVGVVPMSPFGYSALHC 213
DB 535 LRGEQCEVCEVRVLQGLPRE-----YVNARHCLPCHPECPQNGSVT---CFGPEADQC 584
QY 214 -----GILNNIVLCCSYCADLSEIRVRC 237
DB 585 VACAHYKDPDFCVARCPGVKPDLSYMPKFPDDEGACQPCPINCSTHSCVDLDD---KG 641
QY 238 CARRTR-----RLMLRAVRIIAETATMLYSCRTERRRQOF-----IRALLQ 279
DB 642 CPAEQASPLTSIISAVVGVLLVGVVGLIKRRQOKIRKYTMRLRLLQ 692

RESULT 7

W92406
ID W92406 standard; Protein; 1255 AA.
AC W92406;
DT 21-APR-1999 (first entry)
DE Human HER-2/neu oncogene protein.
KW HER-2/neu; oncogene; immune response; T cell; B cell; immunisation;
KW malignancy; treatment; tumour.
OS Homo sapiens.
PH Key Location/Qualifiers
FT Region 576..1255
FT /note= "region which elicits immune response"
FN US5869445-A.
PD 09-FEB-1999.
PF 01-APR-1996; 625101.
PR 01-APR-1996; US-625101.
PR 17-MAR-1993; US-033644.
PR 12-AUG-1993; US-106112.
PR 31-MAR-1995; US-414417.
PR (UNITW) UNIV WASHINGTON.
PI Cheever MA, Disis ML;
DR WPI: 99-152835/13.
DR N-PSDB; X01912.
PT Use of HER-2/neu polypeptides - for eliciting an immune response to an HER-2/neu associated malignancy, particularly for treating or preventing tumours
PS Claim 3; Column 31-38; 26pp; English.
CC This sequence represents the human HER-2/neu oncogene protein. A fragment of this protein is used in a method for eliciting or enhancing an immune

CC response to HER-2/neu protein. The polypeptide can stimulate T cells and B cells to produce an immune response to the HER-2/neu protein. The method can be used for immunisation against a malignancy in which the HER-2/neu oncogene is associated and in the treatment of an existing tumour, or to prevent tumour occurrence or reoccurrence.

CC Sequence 1255 AA;

Query Match 4.6%; Score 74; DB 1; Length 1255;
Best Local Similarity 18.2%; Pred. No. 28;
Matches 64; Conservative 34; Mismatches 117; Indels 136; Gaps 15;

QY 20 SRTPTYSRDLPPPTETTRAT-----ILEDHPLLPCEVTL-----TMHNVSY---VR 63
DB 387 SNTAPLOPEQLQVFTLEETITGLYISAWPDSLPLSVFQNLQVIRGRILHNGAYSLTIQ 446
QY 64 GLPCS-----VGFLLQIE-----WVVPDMLTRELVLIR-----KCMH 98
DB 447 GUGISWLGRLSRLGSLGALIHNNHLCFVHTVPDQLFRNPQALLHTANRDECEVG 506
QY 99 VCLCCANIDIMTSMIHGYESWALCHSCSPGSLQCIAGGQVLASWFRMVVDGAMFNORF 158
DB 507 EGLACHQL-----CARGHCWGPGPTQCVNCSQ-----F 534
QY 159 IWYREVNV-----YMPKEVFMSSVFMGRHLIYLRLWDYGHVGVVPMSPFGYSALHC 213
DB 535 LRGEQCEVCEVRVLQGLPRE-----YVNARHCLPCHPECPQNGSVT---CFGPEADQC 584
QY 214 -----GILNNIVLCCSYCADLSEIRVRC 237
DB 585 VACAHYKDPDFCVARCPGVKPDLSYMPKFPDDEGACQPCPINCSTHSCVDLDD---KG 641
QY 238 CARRTR-----RLMLRAVRIIAETATMLYSCRTERRRQOF-----IRALLQ 279
DB 642 CPAEQASPLTSIISAVVGVLLVGVVGLIKRRQOKIRKYTMRLRLLQ 692

RESULT 8

W03549
ID W03549 standard; Protein; 265 AA.
AC W03549;
DT 21-APR-1997 (first entry)
DE ORF-5 protein sequence from BamHI fragment of HVT.
KW Herpes virus of turkey; open reading frame; ORF; homology; vector;
KW avian herpes virus; recombinant viral vaccine; intergenic region; IBDV;
KW cytomegalovirus immediate early promoter; UL55 gene; repeat region; ILTV;
KW antigen; infectious bursal disease virus; Marek's disease virus; MDV;
KW infectious laryngotracheitis virus; avian anaemia virus; vaccination;
KW infectious bronchitis virus; IBV; poultry; Gumboro disease;
KW Newcastle disease.
PS Herpesvirus of turkey.
PN EP-719864-A2.
PD 03-JUL-1996.
PF 28-DEC-1995; 402970.
PR 30-DEC-1994; FR-016017.
PR (INNR) RHONE MERIEUX SA.
PI Audonnet JCF, Bublout MJM, Dartell RJ, Duinat CV;
PI Laplace ELF, Riviere MAE;
DR WPI: 96-364150/37.
DR N-PSDB; T39309.
PT Live recombinant avian vaccine - comprises herpes virus as vector and having sequence encoding antigenic polypeptide inserted between UL55 gene and repeat region
PS Example 4; Fig 1; 50pp; French.
CC This is the amino acid sequence of the protein encoded by the open reading frame 5 (ORF-5) from the BamHI I fragment from the herpes virus of turkeys (HVT). The DNA fragment was isolated from HVT strain FC126 (J. Gen. Virol., 70:1789-1804 (1989)) and is used as a basis for generating live recombinant viral vaccines using the herpes virus DNA sequences as vectors. The fragment contains 3 intergenic regions into which genes encoding heterologous proteins to be expressed can be inserted, preferably under control of the cytomegalovirus immediate early (CMV-IE) promoter. The heterologous proteins can be used to vaccinate

CC poultry against Gumboro disease (caused by IBV), Newcastle disease,
 CC Marek's disease, infectious bronchitis, infectious laryngotracheitis and
 CC avian anemia.
 SQ Sequence 265 AA;

Query Match 4.6%; Score 73.5; DB 1; Length 265;
 Best Local Similarity 29.4%; Pred. No. 3.8;
 Matches 20; Conservative 9; Mismatches 28; Indels 11; Gaps 2;
 QY 89 ELVILRK--CMHVLCANIDITSMIHGYESWALHCHCSSPGSLQCIAGGOVLASWFR 146
 DB 173 QLMDSKILCLYSACDN-----PGAQGMILSCYGRPGGLQCLAFIRTLMLFN 223
 QY 147 MVVDGAMF 154
 DB 224 DVLSSREF 231

RESULT 9
 W57324
 ID W57324 standard; Protein; 391 AA.
 AC W57324;
 DE Cytosolic glycerol-3-phosphate dehydrogenase GPD1.
 KW Glycerol-3-phosphate dehydrogenase; G3PDH; GPD1; DAR1; OSG1;
 OS Saccharomyces sp.
 PN WO9821340-AL.
 PD 22-MAY-1998.
 PF 10-NOV-1997; U20293.
 PR 13-NOV-1996; US-030602.
 PA (DUPO) DU PONT DE NEMOURS & CO E I.
 PI (GENV) GENENCOR INT INC.
 DR Bulthuis BA, Gatenby AA, Haynie SL, Hsu AK, Lareau RD;
 WI: 98-297943/26.
 N-PSDB; V30822.
 PT Fermentative production of glycerol using recombinant host -
 containing genes for glycerol-3-phosphate dehydrogenase and/or
 glycerol-3-phosphatase
 Claim 9; Page 31-32; 5pp; English.
 CC This Saccharomyces polypeptide comprises a cytosolic
 CC glycerol-3-phosphate dehydrogenase (G3PDH) that catalyses the
 CC conversion of dihydroxyacetone phosphate to glycerol-3-phosphate.
 CC It is encoded by the GPD1 (DAR1, OSG1, D2830, IDL022W) gene (see
 CC V30822). The invention provides recombinant organisms that
 CC express G3PDH and/or glycerol-3-phosphatase (G3P) (see also
 CC W57325-32) useful for the production of glycerol from a variety of
 CC C-sources. A host cell is preferably transformed with a cassette
 CC containing either a G3PDH gene and/or a G3P gene and then cultured
 CC in the presence of a mono-, oligo-, polysaccharide or LC-substrate.
 CC The glycerol obtained is used in cosmetics, liquid soaps,
 CC pharmaceuticals, lubricants and antifreezes; its esters are used in
 CC the oil and fat industries. The method produces glycerol rapidly
 CC and inexpensively without generation of polluting by-products.
 SQ Sequence 391 AA;

Query Match 4.6%; Score 73; DB 1; Length 391;
 Best Local Similarity 23.2%; Pred. No. 7.4;
 Matches 45; Conservative 21; Mismatches 58; Indels 70; Gaps 11;
 QY 45 HPLPE-CNTLTMHNVSVRGLPCSVGFTLIQWVVPDWMVLTREELVILRKMHVCLCC 103
 DB 127 HQFLPRICSQLKHVDHSHVRAISCLKGFEVGAQVQLSSVIT-EELGI-----CCGAL 179
 QY 104 ANIDITSMIHGYESWALHCHCSSPGSLQCIAGGOVLASWFRMVVDGAMFNQFIWYRE 163
 DB 180 SGANIATEV---AQEHWS-----ET 196
 QY 164 VVNNYMPKMEVMSVFMGR-----HLIYLRWYDG---HVGSVVPAMSGYALHCGIL 216
 DB 197 TVAYHIPKD-----FRGEGKVDHVKALFHPYFHV-SVIEDVA-GISI--CGAL 244

QY 217 NNIVLCCSYCADL 230
 DB 245 KNVVALCGGFVEGL 258

RESULT 10
 W60255
 ID W60255 standard; Protein; 391 AA.
 AC W60255;
 DE Klebsiella pneumoniae (first entry)
 KW glycerol-3-phosphate dehydrogenase; production;
 OS 1,3-propanediol; recombinant.
 PN WO9821341-A2.
 PD 22-MAY-1998.
 PF 13-NOV-1997; U20873.
 PR 13-NOV-1996; US-030601.
 PA (GENV) GENENCOR INT INC.
 PI Chase MW, Diaz-torres M, Dunn-coleman NS, Trimbur D;
 DR WI: 98-297944/26.
 N-PSDB; V35737.
 PT New method for increasing production of 1,3-propanediol - comprises
 PT fermentation of inexpensive carbon sources by microorganism
 PT expressing dehydratase, used, e.g. to prolong half-life of enzyme
 PS Disclosure; Page 65-66; 13pp; English.
 CC The sequence is that of cytosolic glycerol-3-phosphate dehydrogenase.
 CC It was used as part of a method of fermentative production
 CC of 1,3-propanediol (1,3-pd), using an organism comprising
 CC at least 1 gene encoding a dehydratase, is improved by
 CC inserting into the host a gene encoding protein X and culturing
 CC the transformant in presence of a carbon source (e.g. mono-, oligo-
 CC or poly-saccharide or LC substrate) convertible to 1,3-pd.
 CC 1,3-pd is a starting material for polyesters, polyurethanes and
 CC cyclic compounds. 1,3-pd can now be produced by a single
 CC recombinant organism from inexpensive carbon sources such
 CC as glucose (rather than costly glycerol or dihydroxyacetone),
 CC rapidly and without causing pollution.
 SQ Sequence 391 AA;

Query Match 4.6%; Score 73; DB 1; Length 391;
 Best Local Similarity 23.2%; Pred. No. 7.4;
 Matches 45; Conservative 21; Mismatches 58; Indels 70; Gaps 11;
 QY 45 HPLPE-CNTLTMHNVSVRGLPCSVGFTLIQWVVPDWMVLTREELVILRKMHVCLCC 103
 DB 127 HQFLPRICSQLKHVDHSHVRAISCLKGFEVGAQVQLSSVIT-EELGI-----CCGAL 179
 QY 104 ANIDITSMIHGYESWALHCHCSSPGSLQCIAGGOVLASWFRMVVDGAMFNQFIWYRE 163
 DB 180 SGANIATEV---AQEHWS-----ET 196
 QY 164 VVNNYMPKMEVMSVFMGR-----HLIYLRWYDG---HVGSVVPAMSGYALHCGIL 216
 DB 197 TVAYHIPKD-----FRGEGKVDHVKALFHPYFHV-SVIEDVA-GISI--CGAL 244
 QY 217 NNIVLCCSYCADL 230
 DB 245 KNVVALCGGFVEGL 258

RESULT 11
 W30680
 ID W30680 standard; Protein; 391 AA.
 AC W30680;
 DE Klebsiella pneumoniae (first entry)
 KW glycerol-3-phosphate dehydrogenase GPD1.
 OS Saccharomyces sp.
 PN WO9821339-A1.
 PD 22-MAY-1998.

10-NOV-1997; U20292.
13-NOV-1996; US-030601.
(DUPO) DU PONT DE NEMOURS & CO E I.
(GEMV) GENENCOR INT INC.
Dias-Forres M, Gatenby AA, Haynie SL, Hsu AK, Lareau RD,
Nagarajan V, Nair RV, Nakamura CE, Payne MS, Picataggio SK,
Trimbur DE, Whited GM;
WPI; 98-297942/26.
N-PSDS; V42016.
Fermentative production of 1,3-propanediol - by single organism
containing cassette comprising specific genes, and capable of using
inexpensive carbon sources
Claim 9: Page 55-56; 95pp; English.
Cytosolic glycerol-3-phosphate dehydrogenase GPD1 is encoded by the
gpd1 gene (see V42016) of Saccharomyces sp. This enzyme catalyses
the conversion of dihydroxyacetone phosphate to glycerol-3-phosphate.
A claimed method for production of 1,3-propanediol (1) comprises
culturing a microorganism transformed with a cassette containing at
least 1 of the genes (see V42012-21) for glycerol-3-phosphate
dehydrogenase, glycerol-3-phosphatase, glycerol dehydratase and
1,3-propanediol oxidoreductase (see W30676-85). Any of these genes
not present on the cassette must be present endogenously. The
enzyme sequences may include substitutions, deletions and additions
provided activity is not altered. A single recombinant organism
can now be used for production of (1) from inexpensive C-sources
(contrast use of glycerol or dihydroxyacetone) without causing
pollution. (1) is a starting material for polyesters, polyurethanes
and cyclic compounds.
SQ Sequence 391 AA;

4.6%; Score 73; DB 1; Length 391;
23.2%; Pred. No. 7.4;
Conservative 21; Mismatches 58; Indels 70; Gaps 11;

45. HPLLPE-CNTLTNNVYVGLPCSVGFTLIQEWVPMWDMVLTREELVILKCMHVCILCC 103
127 HQPLPRICSLKGVDSHVRAISGLKGFVGAQVQLLSYIT-EEELGI-----QCGAL 179
104-ANTDITMSMHIGYESWALCHCHSCPSGLQCIAGGVQLVSWFRMVVDGAMFNQFIWRE 163
180 SGANIATEV--AAQEHWS-----ET 196
164 VVYNNPKVEYFMSSVPMRGR-----HLIYLRLWYDG--HVGSVVPAMSCYSALHCGIL 216
197 TVAIHPKD-----FRGEGKDVRHVLKALFRRFFHV-SVIEDVA-GISI--CGAL 244
217 NNTVVLCCSYCADL 230
245 KNVVLCGGFVEGL 258

RESULT 12
R08222
ID R08222 standard; protein; 624 AA.
AC R08222;
DT 06-MAR-1991 (first entry)
DE Extracellular portion of the human epidermal growth factor receptor 2.
KW Human epidermal growth factor receptor 2; HER2; vaccine; cancer.
OS Homo sapiens.
FH Key
FT Location/Qualifiers
FT 22..32
FT /label= epitope
FT /note= "potential T-cell epitope"
FT 74..84
FT /label= epitope
FT /note= "potential T-cell epitope"
FT 113..130
FT /label= epitope
FT /note= "potential T-cell epitope"
FT 187..197
FT /label= epitope
FT 197..216
FT /label= epitope
FT 216..244
FT /label= epitope

Db		352	LNTNGIEAGVKIRPTEELATADEFRRGYAEYFENKPKPLMHYSVIGFGFDHTKIPN	411
QY		51	CNTLTW-HNVSVRGLPCSGVGFTLQE-----VVVPDWLVLTREE	89
Db		412	GFEWTWFHELEY----PFSRGEVRITSANPYDAPDFDGFNLDERDLMPWYAKKSRET	467
QY		90	LVILRK-----MHVICCANIDIMTSMTHGYESWALHCHC	126
Db		468	ARMESFACEVTSHHPFLFKVDSPAROLDLETCSAYAGPKHLTANTLYHG--SWTPYDK	525
QY		127	SSPGSLQCIAGQGV	140
Db		526	PTPKNDHFVTSNOV	539

RESULT	15
R54843	
ID	R54843 standard; Protein; 1058 AA.
AC	R54843;
DE	HER4 with alternate 3'-end without AP domain.
DT	11-JAN-1995 (first entry)
DE	Human epidermal growth factor receptor; HER; EGF; tyrosine kinase;
KW	cancer; neuronal tissue; muscle tissue; neoplasm;
KW	carcinoma; primer; probe; PCR; autophosphorylation domain; AP.
OS	Homo sapiens.

N-PSDB; Q64898.
New recombinant nucleic acid expressing HER4 - a new receptor tyrosine kinase expressed in some cancer cells, and related vectors, antibodies, ligands etc., for diagnosis and treatment of cancers
Claim 12; Fig 2A; 104pp: English.
The DNA encoding this sequence is identical with that of HER4 (Q64898), up to nucleotide 3168, where the sequence diverges and the ORF stops after 13 amino acids, followed by an extended, unique 3'-UTR.
HER4 is the fourth member of the EGFR-family of tyrosine kinases and is expressed in some human cancers and in some tissues of neuronal or muscle origin. HER4 polynucleotides, opt. labelled, are useful in assays (e.g. of HER4 mRNA) to detect certain neoplasms, esp. breast carcinoma) and as primers in PCR or as probes.
Sequence 1058 AA:
SQ

```
Search completed: June 23, 2000, 10:03:34
Job time: 3362 sec
```


Mon Jun 26 09:10:35 2000

QY 59 VS-----YVGLPCSVGFTLIQWVVPWDMVLTRELVILRKMHCVC--LCCA 104
Db 77 EHVLDNFVSFLEIFHRGDP-SIGRAL--AWMA-W-----CMHACRTLCN 118
QY 105 N-----IDIMTSMI---HGYESWALHCHCSPGSLQCIAGGOVLASWFRMVVDGAMFN 155
Db 119 QSTPYVVDLSVRGMLSEGLDQW-IHQO-----GG-----WSLIEDNIPGS 161
QY 156 QRFWYREVVNMPKEVFMSSVFMGRH 185
Db 162 RRFSTWFLAGLTLSSLVICSYLFSRGRH 191

RESULT 2

PCT-US95-04600-28
; Sequence 28, Application PC/TUS9504600
; GENERAL INFORMATION:
; APPLICANT: LA JOLLA CANCER RESEARCH FOUNDATION
; TITLE OF INVENTION: Interaction of Proteins Involved in
; TITLE OF INVENTION: a Cell Death Pathway
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESS: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/04600
; FILING DATE: 12-APR-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Imbra, Richard J.
; REGISTRATION NUMBER: 37,643
; REFERENCE/DOCKET NUMBER: FP-LJ 1361
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 191 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
PCT-US95-04600-28

Query Match 5.6%; Score 88.5; DB 4; Length 191;
Best Local Similarity 23.3%; Pred. No. 0.013;
Matches 49; Conservative 18; Mismatches 66; Indels 77; Gaps 14;

QY 11 TLRPTRSLRTPYSRDLPPPTETTRATILEDHPLLEPC-----NYLTMHN 58
Db 24 TLHPVLELAARETPL---RLSPEDT---VVLRYHVLEIEIERNSETFTETWNRFIHT 76
QY 59 VS-----YVGLPCSVGFTLIQWVVPWDMVLTRELVILRKMHCVC--LCCA 104
Db 77 EHVLDNFVSFLEIFHRGDP-SIGRAL--AWMA-W-----CMHACRTLCN 118
QY 105 N-----IDIMTSMI---HGYESWALHCHCSPGSLQCIAGGOVLASWFRMVVDGAMFN 155
Db 119 QSTPYVVDLSVRGMLSEGLDQW-IHQO-----GG-----WSLIEDNIPGS 161
QY 156 QRFWYREVVNMPKEVFMSSVFMGRH 185
Db 162 RRFSTWFLAGLTLSSLVICSYLFSRGRH 191

RESULT 3

US-08-467-083-68
; Sequence 68, Application US/08467083
; Patent No. 5726023
; GENERAL INFORMATION:
; APPLICANT: Cheever, Martin A.
; APPLICANT: Disis, Mary L.
; TITLE OF INVENTION: IMMUNE REACTIVITY TO HER-2/NEU PROTEIN
; TITLE OF INVENTION: FOR DIAGNOSIS AND TREATMENT OF MALIGNANCIES IN WHICH
; TITLE OF INVENTION: HER-2/NEU ONCOGENE IS ASSOCIATED
; NUMBER OF SEQUENCES: 68
; CORRESPONDENCE ADDRESS:
; ADDRESS: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: US
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/467,083
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/414,417
; FILING DATE: 06-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Sharkey, Richard G.
; REGISTRATION NUMBER: 32,629
; REFERENCE/DOCKET NUMBER: 920010.448C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERV
; INFORMATION FOR SEQ ID NO: 68:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1255 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-467-083-68

Query Match 5.4%; Score 86; DB 1; Length 1255;
Best Local Similarity 18.5%; Pred. No. 0.42;
Matches 65; Conservative 34; Mismatches 116; Indels 136; Gaps

QY 20 SRRTPYSRDLPPFETETRAT---ILEDHPLLEPCNTL-----TMHNVSY---VR 63
Db 387 SNTAPLOPEQLQVFETLEETGYLYISAWPDSLPDLVSFQNLQVIRGILHNGAYSLTLQ 446
QY 64 GLPCS-----VGFTLIQE-----WVVPWDMVLTRELVILR-----KCMH 98
Db 447 GLGISWGLSLRSLRGSLALIHNTLHCFVHTVPWDQLFRNPQHALLHTANRPEDECVG 506
QY 99 VCLCCANIDIMTSMITHGYESWALHCHCSPGSLQCIAGGOVLASWFRMVVDGAMFNQRF 158
Db 507 EGLACHQL-----CARCHWGPGTQVCNCSQ-----F 534
QY 159 IWYREVVN-----YNNPKEVFMSSVFMGRHLYLYRLMDGHVGSVVPAMSGYSALHC 213
Db 535 LRQECVEECRVLQGLPRE-----YVNARHCLPCHPECPQNGSVT---CFGPEADQC 584
QY 214 -----GILNNVILCCSYCADLSEIRVRC 237
Db 585 VACAHYKDPFCVARGSPGVKPDLSVMPITWKFFDEEGACQCPINCTHSCVDLDD---KG 641
QY 238 CARRTR-----RLMLRAVRIIAETAMLYSCRERRRQQF-----IRALLQ 279
Db 642 CPAEQRASPLTSIISAVVGILLVVLGVVFGILIKRQOKIRKYTMRLLIQ 692

RESULT 5
US-08-486-348A-68
; Sequence 68, Application US/08486348A
; Patent No. 5846538
; GENERAL INFORMATION:
; APPLICANT: Cheever, Martin A.
; APPLICANT: Disis, Mary L.
; TITLE OF INVENTION: IMMUNE REACTIVITY TO HER-2/neu PROTEIN
; TITLE OF INVENTION: FOR DIAGNOSIS AND TREATMENT OF MALIGNANCIES IN WHICH THE
; TITLE OF INVENTION: HER-2/neu ONCOGENE IS ASSOCIATED
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: US
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/486.348A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Sharkey, Richard G.
; REGISTRATION NUMBER: 32,629
; REFERENCE/DOCKET NUMBER: 920010.448C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 68:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1255 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-486-348A-68

Query Match 5.4%; Score 86; DB 2; Length 1255;
Best Local Similarity 18.5%; Pred. No. 0.42;
Matches 65; Conservative 34; Mismatches 116; Indels 136; Gaps 15;
QY 20 SRRTPYSRDLPPPTETTRAT-----ILEDHLLPECNTL-----TMHNVSY---VR 63
Db 387 SNTAPLOEQVFTETLEETGYIYISAWPDSLPLSVFQNLQVIRGRILHNGAYSLTLQ 446
QY 64 GLPCS-----VGFTLIOE-----WVVPDMLTRELVLIR-----KCMH 98
Db 447 GLGISWGLSLRELSGSLALHHNTHLCFVHTVPDOLFPHQALHTANRPEDECVG 506
QY 99 VCLCCANIDIMTSMHIGYESWALHCHSCSPGSLQCIAGGQVLASWFRMVVDGAMFNORF 158
Db 507 EGLACHOL-----CARHCWGPGTQCVCNQ-----F 534
QY 159 IWTREYVN-----YNNPKYMFSSVFMGRHLIYLRWYDGHVGSVVPAMSGYSALHC 213
Db 535 LRQECVEECRVLGLPRE-----YNNARHCLPCHPECPQNGSVT---CFGPEADQC 584
QY 214 -----GILNNIVVLCSCYCADLSEIRVRC 237
Db 585 VACAHYKDPFCVPCVARGKVPDLSPYINWKFPEDEGACQPCINCHSCVDLDD---KG 641
QY 238 CARTR-----RLMLRAVRIIAETATMLYSCRTERROOF-----IRALLQ 279
Db 642 CPAEQRASPLTSIISAVVGLLVVGLVWFGILIKRROQKIRKVTMRRLQ 692
RESULT 6
US-08-468-545B-68

RESULT 4
US-08-414-417B-68
; Sequence 68, Application US/08414417B
; Patent No. 5801005
; GENERAL INFORMATION:
; APPLICANT: Cheever, Martin A.
; APPLICANT: Disis, Mary L.
; TITLE OF INVENTION: IMMUNE REACTIVITY TO HER-2/neu PROTEIN
; TITLE OF INVENTION: FOR DIAGNOSIS AND TREATMENT OF MALIGNANCIES IN WHICH THE
; TITLE OF INVENTION: HER-2/neu ONCOGENE IS ASSOCIATED
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: US
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/414.417B
; FILING DATE: 31-MAR-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Sharkey, Richard G.
; REGISTRATION NUMBER: 32,629
; REFERENCE/DOCKET NUMBER: 920010.448C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 68:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1255 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-414-417B-68

Query Match 5.4%; Score 86; DB 1; Length 1255;
Best Local Similarity 18.5%; Pred. No. 0.42;
Matches 65; Conservative 34; Mismatches 116; Indels 136; Gaps 15;
QY 20 SRRTPYSRDLPPPTETTRAT-----ILEDHLLPECNTL-----TMHNVSY---VR 63
Db 387 SNTAPLOEQVFTETLEETGYIYISAWPDSLPLSVFQNLQVIRGRILHNGAYSLTLQ 446
QY 64 GLPCS-----VGFTLIOE-----WVVPDMLTRELVLIR-----KCMH 98
Db 447 GLGISWGLSLRELSGSLALHHNTHLCFVHTVPDOLFPHQALHTANRPEDECVG 506
QY 99 VCLCCANIDIMTSMHIGYESWALHCHSCSPGSLQCIAGGQVLASWFRMVVDGAMFNORF 158
Db 507 EGLACHOL-----CARHCWGPGTQCVCNQ-----F 534
QY 159 IWTREYVN-----YNNPKYMFSSVFMGRHLIYLRWYDGHVGSVVPAMSGYSALHC 213
Db 535 LRQECVEECRVLGLPRE-----YNNARHCLPCHPECPQNGSVT---CFGPEADQC 584
QY 214 -----GILNNIVVLCSCYCADLSEIRVRC 237
Db 585 VACAHYKDPFCVPCVARGKVPDLSPYINWKFPEDEGACQPCINCHSCVDLDD---KG 641
QY 238 CARTR-----RLMLRAVRIIAETATMLYSCRTERROOF-----IRALLQ 279
Db 642 CPAEQRASPLTSIISAVVGLLVVGLVWFGILIKRROQKIRKVTMRRLQ 692

```

; Sequence 68, Application US/08468545B
; Patent No. 5876712
;
; GENERAL INFORMATION:
; APPLICANT: Cheever, Martin A.
; APPLICANT: Disis, Mary L.
; TITLE OF INVENTION: COMPOUNDS FOR ELICITING OR ENHANCING IMMUNE
; TITLE OF INVENTION: REACTIVITY TO HER-2/neu PROTEIN FOR PREVENTION
; TITLE OF INVENTION: OR TREATMENT OF MALIGNANCIES IN WHICH THE HER-2/neu
; TITLE OF INVENTION: ONCOGENE IS ASSOCIATED
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/625,101
; FILING DATE: 01-APR-1996
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Sharkey, Richard G.
; REGISTRATION NUMBER: 32,629
; REFERENCE/DOCKET NUMBER: 920010.448C7
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1255 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
; US-08-625-101-2

```

```

Query Match 4.6%; Score 74; DB 2; Length 1255;
Best Local Similarity 18.2%; Pred. No. 11;
Matches 64; Conservative 34; Mismatches 117; Indels 136; Gaps 15;

QY 20 SRRTPYSDRLPPETETRAT-----ILEDHPLPECNLT-----TMHNVSY---VR 63
Db 387 SNTAPLOPQLQVFEETLEETGYLYISAWPDSLPLDSVFQNLQVIRGRILHNGAYSLTLQ 446
QY 64 GLPCS-----VGFTLQIE-----WVPDMDVLTREELVILR-----KCMH 98
Db 447 GLGISWGLRLSLRELGLSLALIHNTLCFVHTVPDQLFRNPHQALLHTANRPEDECVG 506
QY 99 VCLCCANIDIMTSMIHGYESWALHCHSCSPGSLQCIAGQVLAWSFRMVVDGAMFNORF 158
Db 507 EGLACHQL-----CARGHCWGPGTQCVNCSQ-----F 534
QY 159 IWTREYVN-----YNNPKYMFMSVFMGRHLYILRLWYDGHVGSVVPAMSGYSALHC 213
Db 535 LRQECVEECRVLQGLPRE-----YVNARHCLPCHPECPQNGSVT---CFGPEADQC 584
QY 214 -----GILNNIVVLCSSYCADLSEIRVRC 237
Db 585 VACAHYKDPFPCVARGCPGVKPDLSYMPIWKFPEDEGACQPCPINCTHSCVDLDD---KG 641
QY 238 CARRTR-----RLMLRAVRIIAETAMLYSCRTERRQOF-----IRALLQ 279
Db 642 CPAEQRASPLTSIISAVVGLLVVGVFGLIKRQOKIRKYTMRLQLQ 692

RESULT 8
US-08-356-786-2
; Sequence 2, Application US/08356786
; Patent No. 5877305

```

```

; Sequence 68, Application US/08468545B
; Patent No. 5876712
;
; GENERAL INFORMATION:
; APPLICANT: Cheever, Martin A.
; APPLICANT: Disis, Mary L.
; TITLE OF INVENTION: IMMUNE REACTIVITY TO HER-2/neu PROTEIN
; TITLE OF INVENTION: FOR DIAGNOSIS AND TREATMENT OF MALIGNANCIES IN WHICH THE
; TITLE OF INVENTION: HER-2/neu ONCOGENE IS ASSOCIATED
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: US
; ZIP: 98104-7092
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/468,545B
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Sharkey, Richard G.
; REGISTRATION NUMBER: 32,629
; REFERENCE/DOCKET NUMBER: 920010.448C5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 68:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1255 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
;
; US-08-468-545B-68

```

```

Query Match 5.4%; Score 86; DB 2; Length 1255;
Best Local Similarity 18.5%; Pred. No. 0.42;
Matches 65; Conservative 34; Mismatches 116; Indels 136; Gaps 15;

QY 20 SRRTPYSDRLPPETETRAT-----ILEDHPLPECNLT-----TMHNVSY---VR 63
Db 387 SNTAPLOPQLQVFEETLEETGYLYISAWPDSLPLDSVFQNLQVIRGRILHNGAYSLTLQ 446
QY 64 GLPCS-----VGFTLQIE-----WVPDMDVLTREELVILR-----KCMH 98
Db 447 GLGISWGLRLSLRELGLSLALIHNTLCFVHTVPDQLFRNPHQALLHTANRPEDECVG 506
QY 99 VCLCCANIDIMTSMIHGYESWALHCHSCSPGSLQCIAGQVLAWSFRMVVDGAMFNORF 158
Db 507 EGLACHQL-----CARGHCWGPGTQCVNCSQ-----F 534
QY 159 IWTREYVN-----YNNPKYMFMSVFMGRHLYILRLWYDGHVGSVVPAMSGYSALHC 213
Db 535 LRQECVEECRVLQGLPRE-----YVNARHCLPCHPECPQNGSVT---CFGPEADQC 584
QY 214 -----GILNNIVVLCSSYCADLSEIRVRC 237
Db 585 VACAHYKDPFPCVARGCPGVKPDLSYMPIWKFPEDEGACQPCPINCTHSCVDLDD---KG 641
QY 238 CARRTR-----RLMLRAVRIIAETAMLYSCRTERRQOF-----IRALLQ 279
Db 642 CPAEQRASPLTSIISAVVGLLVVGVFGLIKRQOKIRKYTMRLQLQ 692

RESULT 7
US-08-625-101-2
; Sequence 2, Application US/08625101
; Patent No. 5869445

```

GENERAL INFORMATION:

APPLICANT: Huston, James S.
APPLICANT: Oppermann, Hermann
APPLICANT: Houston, L. L.
APPLICANT: Ring, David B.
TITLE OF INVENTION: Biosynthetic Binding Protein for Cancer
NUMBER OF INVENTION: Marker
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Edmund R. Pitcher, Testa, Hurwitz, & Thibault
STREET: Exchange Place, 53 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/356,786

FILING DATE:

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/831,967

FILING DATE: 06-FEB-1992

ATTORNEY/AGENT INFORMATION:

NAME: Pitcher, Edmund R.

REGISTRATION NUMBER: 27,829

REFERENCE/DOCKET NUMBER: CRP-053

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 248-7000

TELEFAX: (617) 248-7100

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1255 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-356-786-2

Query Match 4.6%; Score 74; DB 2; Length 1255;
Best Local Similarity 18.2%; Pred. No. 11;
Matches 64; Conservative 34; Mismatches 117; Indels 136; Gaps 15;

QY 20 SRTPTYSRDLPPFETETRAT---ILEDHPLPECNLT---TMHNVSY---VR 63
DB 387 SNTAPLQPEQLQVFTLEETITGLYISAWPDSLPLSVFQNLQVIRGRILHNGAYSILQ 446
QY 64 GLPCS-----VGFLLQE-----WVVPDMVLTREELVILR-----KCMH 98
DB 447 GIGISWGLRSRELGSLALHHTHLCFTVTVPMDQLFRNPHQALLHTANRPEDECVG 506
QY 99 VCLCANTIDMTSMIHGYESWALHCHCSSPGSLQCIAGGQVLASWFRVVDGAMENQRF 158
DB 507 EGLACHQL-----CARGHCWGPQTQCVNSQ-----F 534
QY 159 IYREVVN-----YNNPKEVMEFSSVGRHLIYLRLWDGHVGSVVPAMSGYSALRC 213
DB 535 LRGCVEECRVLOGLPRE-----YVNARHCLPCHPECPQNGSVT---CFGPEADQC 584
QY 214 -----GILNNIVLCCSYCADLSEIRVRC 237
DB 585 VACAHYKDPFCVAPCSVKPFDLSYMPFWKPFDEGACQCPINCHTSCVDLDD---KG 641
QY 238 CARTR-----RLMLRAVRIIAETMTAMLYSCRTERRRQOF-----IRALLQ 279
DB 642 CPASORASPLTSIISAVVGLLVVGLVGVVGLIKRRQOKIRKYTMRLQLQ 692

RESULT 9

US-08-968-563-11
Sequence 11 Application US/08968563
Patent No. 6013494
GENERAL INFORMATION:
APPLICANT: CHARLES E. NAKAMURA
APPLICANT: ANTHONY A. GATENBY
APPLICANT: AMY (KUANG-HUA) HSU
APPLICANT: RICHARD D. LA REAU
APPLICANT: SHARON L. HAYNIE
APPLICANT: MARIA DIAZ-TORRES
APPLICANT: DONALD E. TRIMBUR
APPLICANT: GREGORY M. WHITED
APPLICANT: VASANTHA NAGARAJAN
APPLICANT: MARK S. PAYNE
APPLICANT: STEPHEN K. PICATAGGIO
APPLICANT: RAMESCH V. NAIR
TITLE OF INVENTION: METHOD FOR THE RECOMBINANT
TITLE OF INVENTION: PRODUCTION OF 1,3-PROPANEDIOL
NUMBER OF SEQUENCES: 49
CORRESPONDENCE ADDRESS:
ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
STREET: 1007 MARKET STREET
CITY: WILMINGTON
STATE: DELAWARE
COUNTRY: U.S.A.
ZIP: 19898
ADDRESSEE: GENENCOR INTERNATIONAL, INC.
STREET: 4 CAMBRIDGE PLACE
STREET: 1870 SOUTH WINTON ROAD
CITY: ROCHESTER
STATE: NEW YORK
COUNTRY: U.S.A.
COMPUTER READABLE FORM:
MEDIUM TYPE: 3 50 INCH DISKETTE
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: MICROSOFT WORD FOR WINDOWS 95
SOFTWARE: MICROSOFT WORD VERSION 7.0A
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/968,563
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/030,601
FILING DATE: NOVEMBER 13, 1996
ATTORNEY/AGENT INFORMATION:
NAME: FLOYD, LINDA AXAMETHY
REGISTRATION NUMBER: 33,692
REFERENCE/DOCKET NUMBER: CR-9982
TELECOMMUNICATION INFORMATION:
TELEPHONE: 302-892-8112
TELEFAX: 302-773-0164
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 391 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: GPDI
US-08-968-563-11

Query Match 4.6%; Score 73; DB 3; Length 391;
Best Local Similarity 23.2%; Pred. No. 2.5;
Matches 45; Conservative 21; Mismatches 58; Indels 70; Gaps 11;

QY 45 HPLLE-CNTLTHNVSYVVRGLPCSVGFLIOEWVVPDMVLTREELVILRKCMHVCCLC 103
DB 127 HFLPRICSQLKGHVDHSHVRAISCLAGFEVGARGVOLLSSYIT-EELGI-----QCGAL 179
QY 104 ANIDIMTSMIHGYESWALHCHCSSPGSLQCIAGGQVLASWFRVVDGAMENQRFIYRE 163

Mon Jun 26 09:10:35 2000

Db 180 SGNIATEV---AQEHS-----ET 196
QY 164 VVYNPKVFMFSSVFMGR-----HLIYRLWYDG--HVGSVVPAMSGYSLHGCIL 216
Db 197 TVAXHIPK-----FRGEGRDVKHVKALFHRPYFHV-SVIEDVA-GISI--CGAL 244
QY 217 NNIVLCCSYCADL 230
Db 245 KNYVALGCGFEVL 258

RESULT 10
US-08-991-300-2
; Sequence 2, Application US/08991300
; Patent No. 5973225
; GENERAL INFORMATION:
; APPLICANT: D'VIDIO, RENATO
; APPLICANT: PORCEDDU, ENRICO
; APPLICANT: MERCHITELLI, CINZIA
; APPLICANT: CARDELLI, LUISA ERCOLI
; TITLE OF INVENTION: ISOLATION AND CHARACTERIZATION OF A GENE
; TITLE OF INVENTION: ENCODING A LOW MOLECULAR WEIGHT GLUTENIN
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/991,300
FILING DATE: 16-DEC-1997
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: IT MI 96/A 002663
FILING DATE: 19-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: OBLON, NORMAN F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 2264-0201-0X
TELEPHONE: 703-413-3000
TELEFAX: 703-413-2220
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 369 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-991-300-2

Query Match 4.5%; Score 71; DB 2; Length 369;
Best Local Similarity 21.3%; Pred. No. 4;
Matches 29; Conservative 18; Mismatches 49; Indels 40; Gaps 7;
QY 7 PFGMTLRPTRSR---LSRRTYSDRLPPFTETRTATILEDHPLP-----50
Db 119 PFSQQLPFSQSQPVLPQQPFSQSQPPFSQQLPFSQSQPVLPQQPFSQSQQPPIP 178
QY 50 -----ECMTLMHNVSYVRGL-----PCSVGFTLIQEWVWDM--VLTREELVI 92
Db 179 PQQPPFSQSQQPVLLQQQIPFVHPSTILOQLNPKV---FLQQQCSWPAMPQSLARSMQ 235
QY 93 LRKCMHVC--LCCANI 106

Db 236 QSSC-HVMQOCCOQL 250

RESULT 11
5240838-5
; Patent No. 5240838
; APPLICANT: LEDEBOER, ADRIANUS M.;MAAT, JAN;VERRIPS, CORNELIS P.
; T.VISSER, CHRISTIAAN;JANOWICZ, ZBIGNIEW A.;HOLLEBERG, CORNELIS P.
; TITLE OF INVENTION: REGULATORY SEQUENCES OF ALCOHOL OXIDASE
; (MOX) AND DIHYDROXYACETONESYNTHASE (DAS) OF HANSENULA POLYMORPHA
; NUMBER OF SEQUENCES: 19
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/587,555
; FILING DATE: 24-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 300,211
; FILING DATE: 23-JAN-1989
; APPLICATION NUMBER: 759,315
; FILING DATE: 26-JUL-1985
; SEQ ID NO:5
; LENGTH: 664
5240838-5

Query Match 4.5%; Score 71; DB 5; Length 664;
Best Local Similarity 19.1%; Pred. No. 9.5;
Matches 37; Conservative 24; Mismatches 73; Indels 60; Gaps 7;

QY 1 MTTSGVFGMTLRPTRSR---TPYSRDLPPFTETRTATILE-----DHPLLE 50
Db 352 LITNGIEAGVKIRPTEELATADEDFRGVAEYFENKPKPLMHYSVISGFFGDHTKIPN 411
QY 51 CNTLTM-HNVSYVRGLPCSVGFTLIQ-----WVVPDMVLTREE 89
Db 412 GKFTMTFHELEY---PFSRGFVRIITSANFYADPDPGLNDRDLMPVMYWAYKSKRET 467
QY 90 LVILRKC-----MHVCLCCANIDIMTSMHIGHSWALHCHC 126
Db 468 ARMESFAGEVTSHPFLKVDSPARARDLLETCSAYAGPKHLTANLYHG--SNTVPIDK 525
QY 127 SSPGSLQCIAGQV 140
Db 526 PTPKNDHFVTSNQV 539

RESULT 12
US-08-484-438-4
; Sequence 4, Application US/08484438
; Patent No. 5811098
; Patent No. 5811098 5780031
; GENERAL INFORMATION:
; APPLICANT: Plowman, Gregory D.
; APPLICANT: Culouscou, Jean-Michel
; APPLICANT: Shoyab, Mohammed
; APPLICANT: Siegall, Clay B.
; APPLICANT: Hellstr m, Ingegerd
; APPLICANT: Hellstr m, Karl E.
; TITLE OF INVENTION: HER4 HUMAN RECEPTOR TYROSINE KINASE
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/484,438
FILING DATE: 07-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/323,442
FILING DATE: 14-OCT-1994
APPLICATION NUMBER: US 08/150,704
FILING DATE: 10-NOV-1993
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/981,165
FILING DATE: 24-NOV-1992
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 5624-230
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1058 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-484-438-4

Query Match 4.4%; Score 70.5; DB 2; Length 1058;
Best Local Similarity 24.5%; Pred. No. 21;
Matches 27; Conservative 20; Mismatches 46; Indels 17; Gaps 4;

QY 69 VGFTLQEWVPMWMLTRELVLRLKCMH-----VCLCCANIDIMTMMHGYESW 120
DB 748 VAIKILNETGPKANVEFMDALIMASMDHPLVRLVGCL-SPTQLVTLMPHGCLLE 806
QY 121 ALHCHCSSPGSLQCIAGGQVLASWFMVVDGAMF-NQRFITWYREVYNNM 169
DB 807 YVHEKDNIGS-----QLLNWCVOIAKGMYLEERLVRDLAARNV 849

RESULT 13
US-08-484-438-2
Sequence 2, Application US/08/484438
Patent No. 5811098 5780031
GENERAL INFORMATION:
APPLICANT: Plowman, Gregory D.
APPLICANT: Culouscou, Jean-Michel
APPLICANT: Shoyab, Mohammed
APPLICANT: Siegall, Clay B.
APPLICANT: Hellstr m, Ingegerd
TITLE OF INVENTION: HER4 HUMAN RECEPTOR TYROSINE KINASE
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,438
FILING DATE: 07-JUN-1995
CLASSIFICATION: 530

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/323,442
FILING DATE: 14-OCT-1994
APPLICATION NUMBER: US 08/150,704
FILING DATE: 10-NOV-1993
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/981,165
FILING DATE: 24-NOV-1992
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 5624-230
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1308 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-484-438-2

Query Match 4.4%; Score 70.5; DB 2; Length 1308;
Best Local Similarity 24.5%; Pred. No. 29;
Matches 27; Conservative 20; Mismatches 46; Indels 17; Gaps 4;

QY 69 VGFTLQEWVPMWMLTRELVLRLKCMH-----VCLCCANIDIMTMMHGYESW 120
DB 748 VAIKILNETGPKANVEFMDALIMASMDHPLVRLVGCL-SPTQLVTLMPHGCLLE 806
QY 121 ALHCHCSSPGSLQCIAGGQVLASWFMVVDGAMF-NQRFITWYREVYNNM 169
DB 807 YVHEKDNIGS-----QLLNWCVOIAKGMYLEERLVRDLAARNV 849

RESULT 14
US-08-390-000A-8
Sequence 8, Application US/08390000A
Patent No. 5985583
GENERAL INFORMATION:
APPLICANT: Sealfon, Stuart C.
TITLE OF INVENTION: Cloning and Expression of
TITLE OF INVENTION: Gonadotropin-Releasing Hormone Receptor
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/390,000A
FILING DATE: 17-FEB-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 6923-052
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090
TELEFAX: 212 869-8864/9741
TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:
 LENGTH: 348 amino acids
 TYPE: amino acid
 TOPOLOGY: unknown
 MOLECULE TYPE: protein
 US-08-390-000A-8

Query Match

Best Local Similarity 4.4%; Score 70; DB 2; Length 348;
 Matches 23; Conservative 20; Mismatches 37; Indels 30; Gaps 4;

QY 81 WDMVLTREELVILKCMHVCLCANIDIMTSMIHGYE-SWALHCHCSPGSLQCIAGGQ 139
 DB 126 WSLVLAIERVYV-----VCKPMSNFRFGENHAIMGVAFWVMALACAP----- 171
 QY 140 VLASFWRMVVDG-----AMFNORFIWYREVYNINMPKEYMF 175
 DB 171 PLAGWSRYIPEGLQCSGIDYITLKPEVNNESFVIYMFVYHFTIPMIIF 220

RESULT 15

US-08-422-108-1
 Sequence 1, Application US/08422108
 Patent No. 6015567

GENERAL INFORMATION:

APPLICANT: Hudziak, Robert M.
 APPLICANT: Shepard, H. Michael
 APPLICANT: Ullrich, Axel
 TITLE OF INVENTION: HER2 EXTRACELLULAR DOMAIN
 NUMBER OF SEQUENCES: 2
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Genentech, Inc.
 STREET: 460 Point San Bruno Blvd
 CITY: South San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94080

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: WinPatIn (Genentech)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/422,108
 FILING DATE: 14-Apr-1995
 CLASSIFICATION: 435

PRIOR APPLICATION DATA:

PRIOR APPLICATION NUMBER: 08/355460
 FILING DATE: 13-DEC-1994

PRIOR APPLICATION DATA:

PRIOR APPLICATION NUMBER: 08/048346
 FILING DATE: 15-APR-1993

ATTORNEY/AGENT INFORMATION:

NAME: Lee, Wendy M
 REGISTRATION NUMBER: 00,000

REFERENCE/DOCKET NUMBER:

REFERENCE/DOCKET NUMBER: 554C2D2

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415/225-1994
 TELEFAX: 415/952-9881
 TELEX: 910/371-7168

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
 LENGTH: 624 amino acids
 TYPE: Amino Acid
 TOPOLOGY: Linear

US-08-422-108-1

Query Match 4.4%; Score 70; DB 3; Length 624;
 Best Local Similarity 19.7%; Pred. No. 11;
 Matches 52; Conservative 24; Mismatches 92; Indels 96; Gaps

QY 20 SRRTPYSRDRLPFFETETRAT-----ILEDHPLPECNLT-----TMHNVSY---VR 63
 DB 366 SNTAPLQPEQLQVFETLEEITEYLYISAWPDSLPDLVSFQNLQVIRGRILHNGAISLILQ 425
 QY 64 GLPCS-----VGFTLIQE-----WVVPDMVLTREELVILR-----KCMH 98
 DB 426 GLGISWLGRLSRLGSLALIHHTHLCFVHVTPWDQLFRNPHQALLHTANRPEDECVG 485
 QY 99 VCLCCANIDIMTSMIHGYESWALHCHCSPGSLQCIAGGQVLASFWRMVVDGAMFNORF 158
 DB 486 EGLACHQL-----CARGHCWGPGPTQCVNCSQ-----F 513
 QY 159 IWYREVVN-----YNNPKEVFMSSVEMRGRHLIYLRWDGHVGSVVPAMSFYGSALHC 213
 DB 514 LRQCEVEECRVLOGLPRE-----YVNARHCLPCHPECPQNGSVT---CFGPEADQC 566
 QY 214 GILNNIVVLCSSYCADLSEIRVRC 237
 DB 564 -----VACAHYKDPFPCVARC 579

Search completed: June 23, 2000, 10:04:33
 Job time: 2838 sec

Mon Jun 26 09:10:35 2000

us-09-214-478-2.rai

Page 9

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2000, 09:36:35 ; Search time 61.02 seconds

(without alignments)
282.484 Million cell updates/sec

Title: US-09-214-478-2

Perfect score: 1594

Sequence: 1 MTTSGVPGMTLRPTRSRLS.....RALLQHHRPILMHYDSTPM 294

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 168808 seqs, 58629743 residues

Total number of hits satisfying chosen parameters: 168808

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

PIR_63: *
1: p1r1: *
2: p1r2: *
3: p1r3: *
4: p1r4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1594	100.0	294	1 Q4ADC2	early E4 34K prote
2	858.5	53.9	291	2 S10867	early E4 34K prote
3	331	20.8	153	1 Q4ADB2	early E4 17K prote
4	186.5	11.7	288	1 Q4ADM1	early E4 33K prote
5	88.5	5.6	491	1 Q0B54	BHRF1 protein - hu
6	87	5.5	494	2 T32644	hypothetical prote
7	85.5	5.4	211	1 A70390	formate dehydrogen
8	82	5.1	1603	2 T24098	hypothetical prote
9	81.5	5.1	354	1 T12000	rhodopsin - Africa
10	79.5	5.0	456	2 T16796	hypothetical prote
11	79.5	5.0	562	2 T26242	hypothetical prote
12	79	5.0	352	1 S40688	opsin, rod - Pomat
13	79	5.0	459	2 B58851	NADH dehydrogenase
14	77.5	4.9	378	2 S36506	E2 protein - human
15	77.5	4.9	459	2 T11502	NADH dehydrogenase
16	77.5	4.9	530	2 T30505	hypothetical prote
17	77.5	4.9	642	2 T35003	probable respirato
18	77	4.8	374	2 T05240	pectate lyase (EC
19	77	4.8	446	2 S01187	NADH dehydrogenase
20	76.5	4.8	354	1 A48191	opsin, ultraviolet
21	76.5	4.8	459	1 QXBOAM	NADH dehydrogenase
22	76.5	4.8	459	2 T11411	NADH dehydrogenase
23	76	4.8	505	2 A35993	GTPase (EC 3.1.5.
24	76	4.8	528	2 T05131	hypothetical prote
25	75.5	4.7	373	2 T24487	hypothetical prote
26	75.5	4.7	458	2 S45424	ALG3 protein - yea
27	74.5	4.7	303	2 T32277	hypothetical prote
28	74.5	4.7	321	2 S65225	probable membrane
29	74.5	4.7	895	2 T23191	hypothetical prote
30	74	4.6	196	2 A40623	heat shock protein

ALIGNMENTS

RESULT 1
Q4ADC2
early E4 34K protein - human adenovirus 2
C:Species: Mastadenovirus h2 (human adenovirus 2)
A:Note: Host Homo sapiens (man)
C:Date: 02-Apr-1982 #sequence_revision 02-Apr-1982 #text_change 04-Mar-1994
C:Accession: A03805
R:Herisse, J.; Rigolet, M.; Dupont de Dinechin, S.; Galibert, F.
Nucleic Acids Res. 9, 4023-4042, 1981
A:Title: Nucleotide sequence of adenovirus 2 DNA fragment encoding for the carboxylic
A:Reference number: A93733; M0ID:82059444
A:Accession: A03805
A:Molecule type: DNA
A:Residues: 1-294 <HER>
A:Note: this probable protein was assigned by correlating EM data and S1 digestion st
C:Genetics:
A:Map position: 92.6-95.2
C:Superfamily: adenovirus early E4 34K protein
C:Keywords: early protein

Query Match 100.0%; Score 1594; DB 1; Length 294;
Best Local Similarity 100.0%; Pred. No. 9.3e-142;
Matches 294; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MTTSGVPGMTLRPTRSRLSRRTPYSRDLPPETETETRAILEDHPLPCCNTILTMHNS 60
DB 1 MTTSGVPGMTLRPTRSRLSRRTPYSRDLPPETETETRAILEDHPLPCCNTILTMHNS 60
QY 61 YVRGLPCSVGFTLIQEVVPMVLTREELVILRKCMHVCLCCANIDIMTSMIHYESW 120
DB 61 YVRGLPCSVGFTLIQEVVPMVLTREELVILRKCMHVCLCCANIDIMTSMIHYESW 120
QY 121 ALHCHCSSPGSLQIAGGVVLAWSFRVVDGAMFNQFIWYREVYVNMKEVFMSSVF 180
DB 121 ALHCHCSSPGSLQIAGGVVLAWSFRVVDGAMFNQFIWYREVYVNMKEVFMSSVF 180
QY 181 MRGRHLYLRLWDYGVSVVPMVPMVLTREELVILRKCMHVCLCCANIDIMTSMIHYESW 240
DB 181 MRGRHLYLRLWDYGVSVVPMVPMVLTREELVILRKCMHVCLCCANIDIMTSMIHYESW 240
QY 241 RTRRLMLRAVRITAEETAMLYSCRTERRRQOFIRALLQHHRPILMHYDSTPM 294
DB 241 RTRRLMLRAVRITAEETAMLYSCRTERRRQOFIRALLQHHRPILMHYDSTPM 294

RESULT 2
S10867
early E4 34K protein - human adenovirus 12
C:Species: Mastadenovirus h12 (human adenovirus 12)
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 26-Aug-1999
C:Accession: S10867; S33952
R:Hogenkamp, T.; Esche, H.

Nucleic Acids Res. 18, 3065-3066, 1990

A:Title: Nucleotide sequence of the right 10% of adenovirus type 12 DNA encoding the end

A:Reference number: S10860; MUID:90272430

A:Accession: S10867

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-291 <HOG>

A:Cross-references: EMBL:X51800

R:Sprenkel, J.

submitted to the EMBL Data Library, June 1993

A:Reference number: S33928

A:Accession: S33952

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-227, 'CA', 230-291 <SPR>

A:Cross-references: EMBL:X73487; NID:g313361; PIDN:CAA51901.1; PID:g313386

C:Superfamily: adenovirus early E4 34K protein

Query Match 53.9%; Score 858.5; DB 2; Length 291;
Best Local Similarity 55.0%; Pred. No. 6.2e-73;
Matches 149; Conservative 52; Mismatches 69; Indels 1; Gaps 1;

QY 21 RRPYSDRLPPPTETTRATI-LEDPHLLPECNLTLMHNVYVYVGLPCSVGFTLQEWV 79

Db 10 RLAPYKYQLPCEESKATLSTENSELNPENSLTLHNSEVRGIPSCVGVTLQEWPI 69

QY 80 PWDMLTRELVLKCMHVCCLCANIDIMTSMIHGYESWALHCHCSPGSLQCIAGQ 139

Db 70 PWDMLTDEMFLTKKYSVMCCATINVEVTLQHLGHERWLIHCHCQRPGLSQMSAGM 129

QY 140 VLASFRMVVDGAMFNQRIWREVYVNMPEVMPFMRGRHLIYLRWYDGHVGS 199

Db 130 LLGRWFKMAYGALINKRCFTWREVYVNMPEVMPFMRGRHLIYLRWYDGHVGS 189

QY 200 VYFAMSGYSAHLCGILNNIVLCCSYCADLSEIRVRCARTRRLMLRAVRIIAEETTA 259

Db 190 ALEKYSFGWSAAYGILNNMLVCCDCXDLSEIRMRCPWTRRLMLKVVQVIAENTVR 249

QY 260 MLYSCTERRRQOFTRALQHRPILMHYD 290

Db 250 PLKSHRHERYRQQLKGLIMHRAILFGDYN 280

Query Match 20.8%; Score 331; DB 1; Length 153;
Best Local Similarity 38.5%; Pred. No. 7.2e-24;
Matches 80; Conservative 10; Mismatches 42; Indels 76; Gaps 4;

QY 1 MTTSGVPFGMTLRPTRSLRRTPTYSRDLPPPTETTRATILEDHPHLLPECNLTLMHNV 60

Db 1 MTTSGVPFGMTLRPTRSLRRTPTYSRDLPPPTETTRATILEDHPHLLPECNLTLMHNV 60

QY 61 YVYGLPCSVGFTLQEWVVPWDMVLTRBELYLKCMHVCCLCANIDIMTSMIHGYESW 120

Db 61 YA-----WTSP----- 67

QY 121 ALHCHCSPGSLQCIAGQVGLASW-----FRMVVDGAMFNQRIWREVYV 167

Db 67 -----SPPVEQPOVQGPVQAQLDSDMNLSELPGFEINITDERLARQETVWNTPKM 119

QY 168 NMPKEVMPFMRGRHLIYLRWYD 195

Db 120 SVTHDMLFKA--SRGERTVYVVCWEGG 145

RESULT 4

Q4ADM1

early E4 33K protein - mouse adenovirus 1

C:Species: Mastadenovirus mus1 (mouse adenovirus 1)

A:Note: host Mus musculus (house mouse)

C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 04-Mar-1994

C:Accession: A38519

R:Ball, A.O.; Beard, C.W.; Villegas, P.; Spindler, K.R.

Virology 180, 257-265, 1991

A:Title: Early region 4 sequence and biological comparison of two isolates of mouse a

A:Reference number: A38519; MUID:91082417

A:Accession: A38519

A:Molecule type: DNA

A:Residues: 1-288 <BAL>

A:Cross-references: GB:M37187

C:Superfamily: adenovirus early E4 34K protein

C:Keywords: early protein

Query Match 11.7%; Score 186.5; DB 1; Length 288;
Best Local Similarity 24.8%; Pred. No. 4.7e-10;
Matches 68; Conservative 43; Mismatches 120; Indels 43; Gaps 11;

QY 52 NTLTMHNVYVYVGLPCSVGFTLQEWV-----VPMWMLTRELVLTRK 95

Db 19 NPLTMS--ATLSRMDCCVGLPHCMGWLPSVGAFAVMEGALPIPNYSYDKHVLQ 76

QY 96 CMHVCCLCANIDIMTSM--MIHGYESWALHCHCSPG-----LQCIAGQVGLASWFR 146

Db 77 L--DCLCLFPTCYQSTSYVTGDEVWSLHCHCQROGSLQVGPRLQCLAAAKVRELWVQ 134

QY 147 MVVDGAMFNQRIWREVYVNMPEVMPFMRGRHLIYLRWYDGHVGSVVPAM-- 205

Db 135 KFLIGTRFNEYPOYRVHANRYVNPGLYGVSWC--GKHFYVRA--DGAFFARLKLRA 191

QY 205 SFGYSALHC--GILNNIVLCCSYCA--DLSEIRVRCARTRRLMLRAVRIIAEETAM-- 261

Db 192 RLGGVLFCESSLSCYVYVQCQCACPTDAQVDHCHMRLSLFLLRWQNLGSGSPL 251

QY 261 -----LYSCTERRRQOFTRALQHRPILMHYD 289

Db 252 IPGFDIPRNRTERLRQRMHLRFYSYRTPYRLTY 285

RESULT 5

Q4BE4

BHRE1 protein - human herpesvirus 4 (strain B95-8)

C:Species: human herpesvirus 4, Epstein-Barr virus

C:Date: 25-Feb-1985 #sequence_revision 25-Feb-1985 #text_change 23-Aug-1997

C:Accession: C93065; A03743

R:Bankier, A.T.; Deininger, P.L.; Farrell, P.J.; Barrell, B.G.

Mol. Biol. Med. 1, 21-45, 1983

A:Title: Sequence analysis of the 17,166 bp EcoRI fragment C of B95-8 Epstein-Barr vi

A:Reference number: A93065; MUID:85035713

A:Accession: C93065

A:Molecule type: DNA

A:Residues: 1-191 <BAN>

R:Baer, R.; Bankier, A.T.; Biggin, M.D.; Deininger, P.L.; Farrell, P.J.; Gibson, T.J.; H

Nature 310, 207-211, 1984
A:Title: DNA sequence and expression of the B95-8 Epstein-Barr virus genome.
A:Reference number: A03794; MUID:84270667
A:Contents: annotation; protein coding region
C:Superfamily: human herpesvirus 4 BHRF1 protein
C:Keywords: transmembrane protein

Query Match 5.6%; Score 88.5; DB 1; Length 191;
Best Local Similarity 23.3%; Pred. No. 0.45;
Matches 49; Conservative 18; Mismatches 66; Indels 77; Gaps 14;
QY 11 TLRTSRSLRRTPYSRDLPPFFETRTATILEDHLLPECC-----NLLMHN 58
DB 24 TLHPVLELAARETP-----RLSPEDT-----VVLRYHVLLEELIERNSETFTTWRFTIHT 76
QY 59 VS-----YVGLPCSVGFTLIQWVPMWDMVLTREELVILKCMHVC--LCCA 104
DB 77 EHVLDLDFNSVFLIEFRKGP--SIGRAL--AWNA-W-----CMHACRTLCN 118
QY 105 N-----IDMTSMI-----HGYESWALHCHCSPGSLQIAGGVVLSWFRMVVDGAMFN 155
DB 119 QSTPYVVDLSVRGMEAGLDGW--IHQQ-----GG-----WSTLIEDNIPGS 161
QY 156 QRIWTRVYVNNPYNMREVMSSVFMGRH 185
DB 162 RRFSTWTLFLAGLTLVLVICSYLFSRGRH 191

RESULT 6
T32644
hypothetical protein F56A11.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
A:Reference number: T32644
R:Gatting, S.; Goela, D.; Harper, M.
Submitted to the EMBL Data Library, December 1997
A:Description: The sequence of C. elegans cosmid F56A11.
A:Reference number: 221204
A:Accession: T32644
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-494 <GAT>
A:Cross-references: EMBL:AF038619; PIDN:AAB92076.1; GSPDB:GN00022; CESP:F56A11.4
A:Experimental source: strain Bristol N2; clone F56A11
C:Genetics:
A:Gene: CESP:F56A11.4
A:Map position: 4
A:Introns: 123/3; 235/2; 297/3; 337/3; 401/2; 471/2

Query Match 5.5%; Score 87; DB 2; Length 494;
Best Local Similarity 20.2%; Pred. No. 1.7;
Matches 51; Conservative 40; Mismatches 96; Indels 66; Gaps 10;
QY 36 ETRATILEDHLLPECNITMNVSVRGVGLPCSVGFTLIQWVPMWDMVLTREELVILRK 95
DB 53 EINDTCAEDP---PQADAFRMILIVIGVVCSLGI-----VLTFFLLSLRLDVFERS 103
QY 96 CMHVCLCCANIDMTS-----NMHIG-----YESWALHCHCSPGSLQIAGGQ 139
DB 104 NILYFLACLDLIVELCFMLIFPASLVWDYFRVELLYTCWHFYIKYVS-----TVGQ 156
QY 140 VLASWFRVVDGAMFNQRTWYREVYNNPKEVMS-----SVPMRGHLLYLRLWTD 194
DB 157 VLIAASTLLIVAASFERYICLSKSSIOFSPQRFLISIVGACALPKWSVFEEEL--- 214
QY 195 GHVGSVPANSGFYSAHCGLLNNIVVLCGSCADLSEIRVRCARTRMLMRAVRIIA 254
DB 214 -----QSLPCHPFPQNLRL-----DLSEI-----TRTKYLYIGKLIFS 248
QY 255 EE-TTAMLYSCRT 266

Db 249 ESYKTIWFWCRS 261

RESULT 7
A70390
formate dehydrogenase (EC 1.2.1.2) gamma chain - Aquifex aeolicus
C:Species: Aquifex aeolicus
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: A70390
R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.B.
V. Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MUID:98196666
A:Accession: A70390
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-211 <AAQ>
A:Cross-references: GB:AE000720; NID:92983529; PIDN:AAC07109.1; PID:92983534; GB:AE00
A:Experimental source: strain VF5
C:Genetics:
A:Gene: fdoI
C:Superfamily: formate dehydrogenase gamma chain
C:Keywords: heterotrimer; NAD; oxidoreductase; transmembrane protein

Query Match 5.4%; Score 85.5; DB 1; Length 211;
Best Local Similarity 23.4%; Pred. No. 0.95;
Matches 36; Conservative 20; Mismatches 35; Indels 63; Gaps 7;
QY 64 GLPCSVGFTLIQWVPM--DMVLTREELVILRK-----M 97
DB 67 GVFSIGAVLM---ILKWARDMVITREDIVMLTKIKAYISGRHEELPEVKYNAGKLYA 123
QY 98 HVCCLCANIDMTSMI-----HGYESWALHCH-----CSSPCS 131
DB 124 WVFLSAIVFFLTGLMKWFFENSGMLVRVSVVTHEITFIAGFTIIHYMATVGVPGS 183
QY 132 LQCIAGQVQLASWFRMVVDGAMFNQRIWYREV 165
DB 184 IWGMIGGKVSATW-----AKFHHP-KWYREV 209

RESULT 8
T24098
hypothetical protein R09H10.5 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T24098
R:Basham, V.
Submitted to the EMBL Data Library, July 1996
A:Reference number: Z19840
A:Accession: T24098
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1603 <WII>
A:Cross-references: EMBL:Z77134; PIDN:CAB00874.1; GSPDB:GN00022; CESP:R09H10.5
A:Experimental source: clone R09H10
C:Genetics:
A:Gene: CESP:R09H10.5
A:Map position: 4
A:Introns: 18/1; 57/3; 75/2; 120/1; 168/2; 194/1; 258/1; 302/2; 348/3; 392/3; 442/1;

Query Match 5.1%; Score 82; DB 2; Length 1603;
Best Local Similarity 21.9%; Pred. No. 1.7;
Matches 44; Conservative 29; Mismatches 82; Indels 46; Gaps 11;
QY 9 GMTLRPTRSLRRTPYSRDLPPFFETRTATILEDHLLPECNITMNVSV-----R 63
DB 972 GFSPDPSPDFSRDLTYATHQOPVHL---SSTLQSPDFSDVTSYNNNNIATSLFTR 1028

QY 64 GLPCSVGFTLLQEW-----VVPWMLVRLRELVILKCMHYC-----LCCANIDIM---T 110
Db 1029 TSGKRYETIMNOEYTCOSAGDTFTMETITNTNEVIMRTQRAVCSLDICINGGIMLGN 1088
QY 111 SMIMHGYESWALHCH---CSPGSL-----OC---IAGQVSLASWFMVVDGAMFNORFIW 160
Db 1089 CQCVNGYTS--LHCEVPTCNGSVDFKCCPSIYDG-----DSQYTCNTW 1135
QY 161 YREVVYNNMPKEVFMFSSVPM 181
Db 1136 --NFVETHDPREYNFOQIVFV 1154

RESULT 9
IS1200
rhodopsin - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: IS1200; S31474; S31473
R:Saha, M.S.; Grainger, R.M.
A:Title: Early opsin expression in xenopus embryos precedes photoreceptor differentiation
A:Reference number: IS1200; MUID:93287804
A:Accession: IS1200
A:Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-354 <SAH>
A:Cross-references: GB:S62229; NID:9386187; PID:AB27128.1; PID:9386188; EMBL:L04692; N
R:Knox, B.E.; Scalzetti, L.C.; Batni, S.; Wang, J.Q.
submitted to the EMBL Data Library, December 1992
A:Description: Molecular Cloning of the abundant rod opsin and transducin from Xenopus l
A:Reference number: S31474
A:Accession: S31474
A:Molecule type: mRNA
A:Residues: 1-106, 'Q', 108-136, 'M', 138-240, 'A', 242-354 <KNO>
A:Cross-references: EMBL:L07770; NID:9214734; PID:AA42232.1; PID:9214735
C:Superfamily: vertebrate rhodopsin
C:Keywords: chromoprotein; eye; G protein-coupled receptor; glycoprotein; lipoprotein; R
F:74-96/Domain: transmembrane #status predicted <TM1>
F:93-133/Domain: transmembrane #status predicted <TM2>
F:153-175/Domain: transmembrane #status predicted <TM3>
F:203-230/Domain: transmembrane #status predicted <TM4>
F:253-276/Domain: transmembrane #status predicted <TM5>
F:285-309/Domain: transmembrane #status predicted <TM6>
F:315/Binding site: carbohydrate (Asn) (covalent) #status predicted <TM7>
F:296/Binding site: retinal (Lys) (covalent) #status predicted

Query Match 5.1%; Score 81.5; DB 1; Length 354;
Best Local Similarity 20.0%; Pred. No. 3.9; Mismatches 94; Indels 117; Gaps 13;
Matches 61; Conservative 33;

QY 5 GVPFG-MTLRPTSRSLSRTPYSRDLPPFTETRAILEDHPLPCECNLTLMHNV--- 60
Db 51 GLPINFMLFTVIOHKLRTPLNILL-----NLVFANHEWVLGGFTVMTSHGY 102
QY 60 -----SYVRGLPCSVGFTLLQEWVVPWMLVRLRELVILKCMHYCCLCANIDIMTSM 112
Db 103 FIFGPTGCIIEGF-----FATLGEVALMSLVLAVERYIV-----VCKPMANFRFGNH 152
QY 113 MHGVE-SWALHCHCSFGSLQCIAGQVSLASWFMVVDG-----AMFNOR 157
Db 153 AIMGVATFWINALSCAP-----PLFGNSRYIPEGMCSCGVDDYTLKPENNES 202
QY 158 FIWYREVVYNNMPKEVFMFSSVPMRGRHL----- 187
Db 203 FVIYMFVHTPIPLIVF-----FCYGRLLCTVKEAAQQOESLTTQKAEKVTVMVIMV 258
QY 187 -----IYLRWDVGHVGS-----VVPAMSGYSALHCG-----ILNN-----I 219
Db 259 VFELICWPIYAVFYIFTHQGSNFGVPMVPAFAKSAIYNPVIIVLNKQFNCLI 318

QY 220 VLQCC 224
Db 319 TTLCC 323

RESULT 10

T16796
hypothetical protein T04A6.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 24-Nov-1999
C:Accession: T16796
R:Vaadin, M.; Anderson, K.
submitted to the EMBL Data Library, April 1994
A:Description: The sequence of C. elegans cosmid T04A6.
A:Reference number: Z18579
A:Accession: T16796
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-456 <VAU>
A:Cross-references: EMBL:U00042; NID:9470332; PID:9470335; PIDN:AAA50674.1; CESP:T04A

A:Experimental source: strain Bristol N2

C:Genetics:

A:Gene: CESP:T04A6.3

A:Introns: 23/3; 76/3; 104/3; 232/1; 250/2; 311/2; 340/3; 374/1; 402/3

C:Superfamily: Caenorhabditis elegans hypothetical protein T04A6.3

Query Match 5.0%; Score 79.5; DB 2; Length 456;
Best Local Similarity 19.5%; Pred. No. 7.8; Mismatches 30; Indels 121; Gaps 13;
Matches 54; Conservative 30;

QY 8 FGMLTRPSTRSL-----SRTPYSRDLPPFTETRAILEDHPLPCECNLTLMHNV 60
Db 50 FSVTLNHTPRTALFPWNKSFQTPITAEAP-----ISMLP-CGT----- 89
QY 61 YVRGLPCSVGFTLLQEWVVPWMLVRLRELVILKCMHYCCLCANIDIMTSMTHGYESW 120
Db 89 -----CRYN-----LPVVVIEGKV-----AANF-----SW 111
QY 121 ALHCHCSSPGSLQCTA---GGQVSLASWFMV---VDGAMFNORFIWYREVNV----- 167
Db 112 FNNDCVKSQGLVRLGVPQVWGEYGPPEPANFDAGKIWKPVVNNHKGDISFGV 171
QY 167 YNPKVFMFSSVPMRGRHLIYLRWDVGHVSVVPAMSGY-----SALHCGILNNI 219
Db 172 VNLPCCKQIFAKVDVRNE-----SYIGNQKGEHSARGPACASSY 211

QY 220 VVLCCSYCADL-----SEIRVRCARR 241
Db 212 SILCKLQDNTSRDRKNTYSVLSDDDNDVRLNCGKR 248

RESULT 11

T26242
hypothetical protein W06F12.2a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T26242
R:Lloyd, C.
submitted to the EMBL Data Library, December 1996
A:Reference number: Z20179
A:Accession: T26242
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-562 <WIL>
A:Cross-references: EMBL:Z83244; PIDN:CAB54314.1; GSPDB:GN00021; CESP:W06F12.2a
A:Experimental source: clone W06F12
C:Genetics:
A:Gene: CESP:W06F12.2a
A:Map position: 3
A:Introns: 23/3; 75/1; 153/3; 201/1; 283/1; 305/3; 332/2; 555/3

us-09-214-478-2.rpr

Mon Jun 26 09:10:37 2000

C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 07-Dec-1999

C:Accession: B58851

R:Aranson, U.; Gullberg, A.; Widegren, B.

J. Mol. Evol. 33, 556-568, 1991

A:Title: The complete nucleotide sequence of the mitochondrial DNA of the fin whale.

A:Reference number: A58850; MUID:92139449

A:Accession: B58851

A:Molecule type: DNA

A:Residues: 1-459 CARN

A:Cross-references: GB:X61145; NID:g12772; PIDN:CAA43448.1; PID:g12780

C:Genetics:

A:Genome: mitochondrion

A:Genetic code: SGCI

A:Superfamily: NADH dehydrogenase (ubiquinone) chain 4

C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation

Query Match 5.0%; Score 79; DB 2; Length 459;

Best Local Similarity 19.5%; Pred. No. 8.8; Indels 74; Gaps 9;

Matches 37; Conservative 29; Mismatches 50;

QY 81 WDMVLR-----EELVLRKCMHVCCLCCANIDIMTS-----MMIHGYESMAL 122

DB 267 WGMIMTSSICLRQDLKSLIAYSSVSHMALVIAAILIQTWPMSYGATAMIAHGLTSSML 326

QY 123 RHCSS-----PGLQCIAGGOVLASWFRVVDGAMFNORFIWYREVNNMP 170

DB 327 FCLANSNYRIHSRTMILPGLQVFL--PLMASWLLA-----SLTNLALP 370

QY 171 K-----EVMFMSSVF-----MRGRHLIYLRM-----YGHVGVSVVPAM 204

DB 371 PTNLIGELVVMVSFNSNTILLMGINIVITALYSLYMLTQKTHHNNITPSF 430

QY 205 SFGYS--ALH 212

DB 431 TREHALMALH 440

RESULT 14

S36506

E2 protein - human papillomavirus type 30

C:Species: human papillomavirus type 30

C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 26-Aug-1999

C:Accession: S36506

R:Delius, H.; Hofmann, B.

Submitted to the EMBL Data Library, August 1993

A:Description: Primer-directed sequencing of human papillomavirus types.

A:Reference number: S36469

A:Accession: S36506

A:Molecule type: DNA

A:Residues: 1-378

A:Cross-references: EMBL:X74474; NID:g396973; PIDN:CAA52546.1; PID:g396977

A:Superfamily: papillomavirus E2 protein

C:Keywords: DNA binding; early protein; transcription regulation

Query Match 4.9%; Score 78; DB 2; Length 378;

Best Local Similarity 21.6%; Pred. No. 8.9; Indels 68; Gaps 14;

Matches 50; Conservative 27; Mismatches 86;

QY 74 IOEYVVPDMLTRELVLRLK-----CMHVCLCCANIDIMTSMMI----- 115

DB 26 IEDHYVYKAV--RHENVLYKARQNNITKLHQQVPCQVCKAKACVAIEIQMALESLEY 83

QY 115 ---HGYESWALHCHCSS---PGSLQCI-AGGOVLASWFRVVDGAMFN--ORFIWYREV 165

DB 84 KTEYKVEEWTLDVCENMMHTAPKQCFKSKGRIVWF-----DGKDNREYVVM--QWV 137

QY 166 NY-----NMPKEYMFMSSVEMRGRHLIYLRWYDGHVGVSVVPAMSGYSA---LHCGI 215

DB 138 YCGDNGWTKVPSVDYKGIYYVHDGNKYVYTDND-----EAVKYGKGTWVHMG- 190

QY 216 LNNIVLCCSYCADLSIRVRCARRRRLMLRAVRIIAEETAMLYSRT 266

Query Match 5.0%; Score 79.5; DB 2; Length 562;

Best Local Similarity 23.4%; Pred. No. 9.8; Indels 55; Gaps 9;

Matches 39; Conservative 25; Mismatches 48;

QY 100 CLCCANID-----IMTSMIHGYESWALHCHC---SPGSLQCC----- 135

DB 6 CCCSSKDAITIGISLVYALASLLFGWQVGLN-HCRVVTMAQSNLQCEWDCPCVGA 64

QY 135 -----IAGG-----OVL-----ASWFRVVDGAMFNORFIWYREVNNMPKVEFMSSVF 180

DB 65 SSKRTSALIEGLFVQVLCIAFFLVSASMIYGIHTWSKYL---WPFVFFVLMSSIL 121

QY 181 MRGRHLIYLRWYDGHVGVSVVPAMSGYSAHCGILNNIVLCCSYC 227

DB 122 AT---LAYCIMWCGDVRVYLAIT-----IIEIIVFINIYC 156

RESULT 12

S40688

opsin, rod - Pomatoschistus minutus

C:Species: Pomatoschistus minutus

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

C:Accession: S40688

R:Archer, S.N.; Lythgoe, J.N.; Hall, L.

Proc. R. Soc. Lond. B Biol. Sci. 248, 19-25, 1992

A:Title: Rod opsin cDNA sequence from the sand goby (Pomatoschistus minutus) compared with other teleost sequences

A:Reference number: S40688; MUID:92396703

A:Accession: S40688

A:Molecule type: mRNA

A:Residues: 1-352 CARC

A:Cross-references: EMBL:X62405; NID:g433817; PIDN:CAA44275.1; PID:g433818

C:Superfamily: vertebrate rhodopsin

C:Keywords: chromoprotein; eye; G protein-coupled receptor; glycoprotein; lipoprotein; F; 37-61/Domain: transmembrane #status predicted <TM1>

F; 74-96/Domain: transmembrane #status predicted <TM2>

F; 114-140/Domain: transmembrane #status predicted <TM3>

F; 153-173/Domain: transmembrane #status predicted <TM4>

F; 203-230/Domain: transmembrane #status predicted <TM5>

F; 233-276/Domain: transmembrane #status predicted <TM6>

F; 286-309/Domain: transmembrane #status predicted <TM7>

F; 110-187/Disulfide bonds: #status predicted

F; 296/Binding site: retinal (Lys) (covalent) #status predicted

Query Match 5.0%; Score 79; DB 1; Length 352;

Best Local Similarity 19.9%; Pred. No. 6.6; Indels 98; Gaps 13;

Matches 51; Conservative 31; Mismatches 76;

QY 55 TMHNVSYRGLPCSV-GP--TLIOEYVVPDMLTRELVLRLKCMHVCCLCCANIDIMTS 111

DB 98 SMHGTVFLGRGCVGEGFATLGG-IALMSLVLAVERVWV-----VCKPISNFRFTEN 151

QY 112 MHIHGYE-SWALHCHCSSPGSLQCIAGGOVLASWFRVVDG-----AMFNQ 156

DB 152 HAIMGVAFSWIAATCAVP-----PLVGMRSRYIPEGMCQSCGVDTYTRAEGENNE 201

QY 157 RFTWYREVNNMPKVEFMSSVEMRGRHLIYLR-----FCVGRLLCAVKEAQAQSEETQRAERETVRMVIIM 257

DB 202 SFVIYMFIVHFLAPLVIF-----FCVGRLLCAVKEAQAQSEETQRAERETVRMVIIM 257

QY 191 -----LWV-----DGHVGVSVVPAMSGYSAHCGILNNIVLCCS--- 226

DB 258 VIGFLSWLPYASVAVYTHQGTGFLPMTIPAFKSSALY-----NPMIYICMKNQF 313

QY 226 -YCADLSEIRVRCAR 240

DB 314 RHCM-----ITTLACGK 325

RESULT 13

B58851

NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 4 - finback whale mitochondrion

C:Species: mitochondrion Balaenoptera physalus (finback whale, common rorqual)

Db 190 NESI-----YCPD-----SVSILRSNVPVETWYNTYNYQPT 226

RESULT 15

TI1502

NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 4 - dog mitochondrion (fragment)

C:Species: mitochondrion Canis lupus familiaris (dog)

C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 07-Dec-1999

C:Accession: T11502

R:Kim, K.S.; Lee, S.E.; Jeong, H.W.; Ha, J.H.

Mol.:Phylogenet. Evol. 10, 210-220, 1998

A:Title: The complete nucleotide sequence of the domestic dog (Canis familiaris) mitochondrion

A:Reference number: Z17276; MUID:99097286

A:Accession: T11502

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-459 <KIM>

A:Cross-references: EMBL:U96639; NID:g4154170; PID:g4154180; PIDN:AAD04772.1

C:Genetics:

A:Genome: mitochondrion

A:Note: ND4

C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;

Query Match 4.9%; Score 77.5; DB 2; Length 459;

Best Local Similarity 24.1%; Pred. No. 12;

Matches 26; Conservative 14; Mismatches 37; Indels 31; Gaps 5;

QY 40 TILEDHPLLPCEWTLMHNSVYVGRGLPCSVGFTLIQEWVYP-----WDMVLTRRELVLK 95

Db 153 TLMGSLPLIVA-----LLYHFMGSLNFMIOYWIQIPLNSWSNIF----- 195

QY 96 CMHVCCLCANTIDMTSMIRHGYESWALHCHCSFSGSLQCIAGQVLA 143

Db 195 ---LWLACM-NAFVNMPLYGLHLWLPKAVHAP-----IAGSMVLAA 233

Search completed: June 23, 2000, 10:05:43

Job time: 1748 sec

Mon Jun 26 09:10:39 2000

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2000, 10:03:39 ; Search time 13.99 Seconds
(without alignments)
640.012 Million cell updates/sec

Title: US-09-214-478-2
Perfect score: 1594
Sequence: 1 MTSGVPGMTLRTRSLRSL.....RALLQHRPILMHYDSTPM 294
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 83857 seqs, 30454973 residues
Total number of hits satisfying chosen parameters: 83857
Minimum DB seq length: 0
Maximum DB seq length: 1000000
Post-processing: Minimum Match 0%
Listing first 45 summaries
Database : SwissProt_38.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	Query Match	Length	ID	Description
1	1594	100.0	294	1 E434_ADE02	P03239 human adeno
2	990	62.1	292	1 E434_ADE09	P89083 human adeno
3	874.5	54.9	291	1 E434_ADE12	P36710 human adeno
4	706	44.3	257	1 E434_ADE40	P64865 human adeno
5	331	20.8	153	1 E417_ADE02	P03238 human adeno
6	253.5	15.9	259	1 E434_ADECT	P87568 canine adeno
7	252.5	15.8	265	1 E434_ADECR	P96690 canine adeno
8	250.5	15.7	265	1 E434_ADECC	P65962 canine adeno
9	186.5	11.7	289	1 E434_ADEM1	P23125 mouse adeno
10	91.5	5.7	643	1 UROM_BOVIN	P48733 bos taurus
11	88.5	5.6	131	1 FAR_EBV	P03182 epstein-bar
12	81.5	5.1	334	1 OPSP_XENLA	P29403 xenopus lae
13	79	5.0	289	1 OPSP_LIMBE	O42427 limocottus
14	79	5.0	352	1 OPSP_POMMI	P35403 ponatoschis
15	79	5.0	354	1 OPSP_ZEUFA	O42604 zeus faber
16	78	5.0	459	1 NU4M_BALPH	P24975 balaenopter
17	78	4.9	378	1 VE2_HPV30	P36790 human papil
18	77.5	4.9	459	1 NU4M_CANFA	P35350 brachydanio
19	77	4.8	289	1 OPSP_BATMU	P03910 bos taurus
20	76.5	4.8	354	1 OPSP_BRARE	P48016 felis silve
21	76.5	4.8	459	1 NU4M_BOVIN	P48016 felis silve
22	76.5	4.8	459	1 NU4M_FELCA	P48016 felis silve
23	76.5	4.8	598	1 NTG1_MUSCO	P48016 felis silve
24	76	4.8	504	1 DQTP_ECCLI	P15723 escherichia
25	76	4.8	504	1 DQTP_ECCLI	P15723 escherichia
26	76	4.8	504	1 DQTP_ECCLI	P15723 escherichia
27	75.5	4.7	459	1 PELL_YEAST	P25578 saccharomyc
28	75	4.7	349	1 ALG3_YEAST	P38179 saccharomyc
29	74	4.6	196	1 HTGA_ECCLI	P79798 myrpristis
30	74	4.6	350	1 IL8A_GORGO	P28697 escherichia
31	74	4.6	1255	1 ERB2_HUMAN	P55919 gorilla gor
32	73.5	4.6	446	1 NU4M_DROYA	P04626 homo sapien
33	73.5	4.6	475	1 YKGF_ECCLI	P07707 drosophila
34	73.5	4.6	604	1 KRAA_RAT	P77536 escherichia
					P14056 rattus norv

35	73.5	4.6	606	1 KRAA_HUMAN	P10398 homo sapien
36	73.5	4.6	606	1 KRAA_PIG	O19004 sus scrofa
37	73	4.6	289	1 OPSP_PARKN	O42452 paracottus
38	73	4.6	348	1 OPSP_TRIMA	O62796 trichechus
39	73	4.6	391	1 GPD1_YEAST	Q00553 saccharomyc
40	73	4.6	504	1 GTP_SHIBO	Q59827 shigella bo
41	73	4.6	515	1 GLCM_MOUSE	P17439 mus musculu
42	72.5	4.5	352	1 C5AR_CANFA	P30992 canis famli
43	72.5	4.5	459	1 NU4M_BALMU	P41298 balaenopter
44	72.5	4.5	459	1 NU4M_SHEEP	O78755 ovis aries
45	72	4.5	289	1 OPSP_ABYKO	O42294 abyssocottu

ALIGNMENTS

RESULT 1

E434_ADE02 STANDARD; PRT; 294 AA.

AC P03239;

DT 21-JUL-1986 (Rel. 01, Created)

DT 21-JUL-1986 (Rel. 01, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE EARLY E4 34 KD PROTEIN.

OS Human adenovirus type 2.

OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Masgadenovirus.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE; 82059444.

RA Herisse J., Rigolet M., Dupont de Dinechin S., Galibert F.;

RT "Nucleotide sequence of adenovirus 2 DNA fragment encoding for the carboxylic region of the fiber protein and the entire E4 region.";

RL Nucleic Acids Res. 9:4023-4042(1981);

CC -!- MISCELLANEOUS: FOUND IN ASSOCIATION WITH E1B 55 KD PROTEIN

CC -!- SIMILARITY: BELONGS TO THE ADENOVIRUS E4 30 TO 34 KD PROTEIN FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

CC -----

CC EMBL; J01917; -; NOT_ANNOTATED_CDS.

DR PIR; A03805; Q4ADC2.

KW Early protein.

SQ SEQUENCE 294 AA; 34116 MW; F60C83A38240BE0C CRC64;

Query Match 100.0%; Score 1594; DB 1; Length 294;

Best Local Similarity 100.0%; Pred. No. 6.3e-148;

Matches 294; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MTSGVPGMTLRTRSLRSLRRTPSYSDRLPPPTETETRAILEDHPLLPNCNTLTHMNV 60

Db 1 MTSGVPGMTLRTRSLRSLRRTPSYSDRLPPPTETETRAILEDHPLLPNCNTLTHMNV 60

Qy 61 YVRLPCSVGTLTQEWVVPDWMVLTREELVILKRMHVCLCCANIDIMTSMHIGYESW 120

Db 61 YVRLPCSVGTLTQEWVVPDWMVLTREELVILKRMHVCLCCANIDIMTSMHIGYESW 120

Qy 121 ALHCHCSPGSLQCIAGGQVLASFRMVDGMNQFRTWTVREVNTNMPKVMFMSV 180

Db 121 ALHCHCSPGSLQCIAGGQVLASFRMVDGMNQFRTWTVREVNTNMPKVMFMSV 180

Qy 181 MRGRHLIYLRLWDHGVSVVPAMFGYSALHCGTLNNIVLVCCSYCADLSEIRVRCAR 240

Db 181 MRGRHLIYLRLWDHGVSVVPAMFGYSALHCGTLNNIVLVCCSYCADLSEIRVRCAR 240

Qy 241 RTRELMLRAVRIIAETTTAMLYSCRTRRROOFIRALLQHRPILMHYDSTPM 294

Db 241 RTRMLRAVRIIAETAMLYSCRTERRRQOFIRALLOHHRPILMHVDYDSTPM 294

RESULT 2
E434_ADE09
ID E434_ADE09 STANDARD; PRT; 292 AA.
AC P89083;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE EARLY E4 34 KD PROTEIN.
OS Human adenovirus type 9.
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 96327849.
RA Javlier R., Shenk T.;
RT "Mammary tumors induced by human adenovirus type 9: a role for the
RL viral early region 4 gene";
RL Breast Cancer Res. Treat. 39:57-67(1996).
CC -!- SIMILARITY: BELONGS TO THE ADENOVIRUS E4 30 TO 34 KD PROTEIN
CC FAMILY
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL; S82508; AAB37508.1;
KW Early protein.
SQ SEQUENCE 292 AA; 34082 MW; 1F0815EDEF2A2A1E CRC64;
Query Match 62.1%; Score 990; DB 1; Length 292;
Best Local Similarity 63.7%; Pred. No. 3.8e-89;
Matches 174; Conservative 45; Mismatches 52; Indels 1; Gaps 1;
QY 21 RRTPYSRDLPPPTETRAILEDHPLPECNLTAMHNSYVGRGLPCSVGFTLIQEWVVP 80
Db 11 RHPYRRLPRSDETRSLAEQHPFLPCDHDYHNSVSSVGRGLPCAAGFTLLQEPFVP 70
QY 81 WDMVLTRRELVLKRCMHVCLCCANIDIMTSMIHGYESWALHCHCSPGSLQCIAGSQV 140
Db 71 WDMILTPPEIKILKRCMSYCLCPATLDLVRAQMSVGYERWILHCHCSPGSLQCRAGGTL 130
QY 141 LASWFRMVVDGAMFNQRFIWRVNVNPKVFMVMSVFMGRHLLIYLRWYDGHVGSV 200
Db 131 LAVWFRVYVGMFNQRFIWRVNVNPKVFMVMSVFMGRHLLIYLRWYDGHVGSV 190
QY 201 VPAMSGYSAHCGILNNIVVLCSSYCADLSEIRVCCARRTRRLMLRAVRIIAET--T 258
Db 191 IPNMSFGWSALNYGLNNVIMWCCTICENNAEIRMRCCARRTRRLMLKAVGIIVRETCDP 250
QY 259 AMLYSCRTERRRQOFIRALLOHHRPILMHVDYD 291
Db 251 DPICSTRTEPRQRLRALMRHRPILFSEYES 283

RESULT 3
E434_ADE12
ID E434_ADE12 STANDARD; PRT; 291 AA.
AC P36710;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE EARLY E4 34 KD PROTEIN.
OS Human adenovirus type 12.
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
RN [1]
RP SEQUENCE FROM N.A.

Db 241 RTRMLRAVRIIAETAMLYSCRTERRRQOFIRALLOHHRPILMHVDYDSTPM 294

RESULT 4
E434_ADE40
ID E434_ADE40 STANDARD; PRT; 257 AA.
AC Q64865;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE EARLY E4 30 KD PROTEIN.
OS Human adenovirus type 40.
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DUGAN;
RX MEDLINE; 94087748.
RA Davison A.J., Telford E.A., Watson M.S., McBride K., Mautner V.;
RT "The DNA sequence of adenovirus type 40.";

RX MEDLINE; 94076430.
RA Sprengel J., Schmitz B., Heuss-Neitzel D., Zock C., Doerfler W.;
RT "Nucleotide sequence of human adenovirus type 12 DNA: comparative
RL functional analysis";
RL J. Virol. 68:379-389(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 90272430.
RT Hogenkamp T., Esche H.;
RT "Nucleotide sequence of the right 10% of adenovirus type 12 DNA
RT encoding the entire region E4";
RL Nucleic Acids Res. 18:3065-3066(1990).
CC -!- SIMILARITY: BELONGS TO THE ADENOVIRUS E4 30 TO 34 KD PROTEIN
CC FAMILY
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL; X73487; CAAS1901.1;
KW Early protein.
SQ SEQUENCE 291 AA; 34383 MW; CDIA07D37FBB5E96 CRC64;
Query Match 54.9%; Score 874.5; DB 1; Length 291;
Best Local Similarity 55.7%; Pred. No. 6.5e-78;
Matches 151; Conservative 52; Mismatches 67; Indels 1; Gaps 1;
QY 21 RRTPYSRDLPPPTETRAILEDHPLPECNLTAMHNSYVGRGLPCSVGFTLIQEWVVP 79
Db 10 RLAPYKYLPPCEQSKATSTSENSLPWCNLSLTHNVSEVRGIPSCYFTVLQEWPI 69
QY 80 PDMVLTRRELVLKRCMHVCLCCANIDIMTSMIHGYESWALHCHCSPGSLQCIAGSQ 139
Db 70 PDMILTPPEIKILKRCMSYCLCPATLDLVRAQMSVGYERWILHCHCSPGSLQCRAGGTL 129
QY 140 VASWFRMVVDGAMFNQRFIWRVNVNPKVFMVMSVFMGRHLLIYLRWYDGHVGSV 199
Db 130 LLGRWFKNAVYALINKRCFWYREVVAHLPKPYVYGVSTFVRGRHLLYFKIMYDGHAWL 189
QY 200 VVPAMSGYSAHCGILNNIVVLCSSYCADLSEIRVCCARRTRRLMLRAVRIIAETTA 259
Db 190 ALEKVSFGWSAFNYGLNNVIMWCCTICENNAEIRMRCCARRTRRLMLKAVGIIVRETCDP 249
QY 250 MLYSCRTERRRQOFIRALLOHHRPILMHVDYD 290
Db 250 PLKSRHRYRQQLKGLIMHRAILFGDYN 280

```

DR TRANSFAC; T01301; -.          EARLY 17 KD PROTEIN.
KW EARTY protein.                EARLY 10 KD PROTEIN.
FT CHAIN             56         153
SQ SEQUENCE        153 AA; 17404 MW; 1B3A657DAD8557CC CRC64;

Query Match               20.8%; Score 331; DB 1; Length 153;
Best Local Similarity    38.5%; Pred. No. 2,4e-25;
Matches      80; Conservative   10; Mismatches     42; Indels    76; Gaps      4;

QY      1 MTTSVPPGMLRPRSLSRREYSRDLRPPEETERRATILEDHPPLPCNTLTNHNVS 60
       1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      1 MTTSVPFGMLRLPFRSLSRRTSPRSRRLPPEETERRAILEDHPLPCNTLTNHNVS 60
       1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY      61 YVRGLPCSGETLLDEWVPMDNYLTRBELYLKRCMNVCLCCANIDMTSMIHGYESW 120
       1 | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
DB      61 YA-----WTSP-----                          ----- 67

QY      121 ALHCSSPGSLQCIAGQVLASW-----FRMVGDGMFNORFIYREVYN 167
       1 | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
DB      67 -----SPPVBPPOVGGOVAQQQLDSMNLSLSELPGEFINITDERLARQEIWNITPKM 119
       1 | | | | | : | : | : | : | : | : | : | : | : | : | : | : |

QY      168 NMPKEVFMSVFMGRHLIYLRYDG 195
       ::::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB      120 SYTHDMKLTK--SRGERITYSVCEGG 145

RESULT      6
E434_ADECT ID E434_ADECT STANDARD: PRT; 259 AA.
AC p87568;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE EARLY E4 30 KD PROTEIN.
NC Canine adenovirus type 2 (strain Toronto A 26-61)
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
[1]
RP SEQUENCE FROM N.A.
RA Campbell J B., Zhao Y.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DDBJ databases.
CC -1 SIMILARITY: BELONGS TO THE ADENOVIRUS E4 30 TO 34 KD PROTEIN FAMILY.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sdb.ch/announce/or send an email to license@isb-sdb.ch).
CC -----
DR EMBL; U77082; AAB38735.1; -.
KW Early Protein.
SQ SEQUENCE 259 AA; 30014 MW; 9C966CA011C2A745 CRC64;

Query Match              15.9%; Score 253.5; DB 1; Length 259;
Best Local Similarity    26.3%; Pred. No. 1.5e-17;
Matches      70; Conservative   47; Mismatches     86; Indels    63; Gaps      9;

QY      51 CNTLTNMHVSYVRGLP---CSVGFELLIDEWVPMDNYLTR-----ELVIIR 94
       1 | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
DB      5 CAAETTCVAVTAIVRKVCNC-FALCEIPFLIMDVIAHRHKLLFGSGFCMGAEIIIVTK 63
       1 | | | | | : | : | : | : | : | : | : | : | : | : | : | : |

QY      95 KCMHYCLCCANIDMTSMIHGYESMALHCHCSSPSGLQCIAGQVLASFRRMYVDGMF 154
       1 | | | | | : | : | | | | | : | : | : | : | : | : | : | : |
DB      64 HC---CLAEA-----OTWRHCHCSSNSLSGCCMAKRHYVKVEDIFIKGAM 107
       1 | | | | | : | : | : | : | : | : | : | : | : | : | : | : |

QY      155 NORFIYRVVVVNNMKREMVMEMSGRLIYLRLMW-----DGHVG 198
       1 | | | | | | | | | | | : | : | : | : | : | : | : | : | : |
DB      108 NRKYMMVREFVNSSRPDELINTVGSIIFNTNTYIIFRLSFPTVAKAMEAFKKRCINELG 167

```

QY 199 SVPMASFGSALHGLINNIIVLCCSYCADLSEIRVRCARTRRLMRAVIAEETT 258
 DB 168 VLKSC-----TYNWLVLKCKSCSLQNYCALKNAWVSVIYVAKVEY-ETKP 215
 QY 259 AMY--SCRTERRRROQTRIRALLQHR 282
 DB 216 VILHQTTSKAERRRQSLRQAMMGR 241
 RESULT 7
 E434_ADECR STANDARD: PRT: 265 AA.
 AC 096690;
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DE EARLY E4 31 KD PROTEIN.
 OS Canine adenovirus type 1 (strain RI261).
 OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 97275900.
 RA Morrison M.D., Orlowski D.E., Nicolson L.;
 RT "Complete DNA sequence of canine adenovirus type 1."
 RL J. Gen. Virol. 78:873-878(1997).
 CC -1- SIMILARITY: BELONGS TO THE ADENOVIRUS E4 31 TO 34 KD PROTEIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; Y07760; CAA69046.1; -
 KW Early protein.
 SQ SEQUENCE 265 AA; 30737 MW; 14B3DC9DCB1D2E15 CRC64;
 Query Match 15.88; Score 252.5; DB 1; Length 265;
 Best Local Similarity 26.98; Pred. No. 2e-17;
 Matches 70; Conservative 43; Mismatches 96; Indels 51; Gaps 8;
 QY 51 CNTLMHNVSYVGRGP--CSVGFLLIOEWVPMVMYTRRELYI-----LRKM 97
 DB 11 CNAETSHVAVVAPAFPCNC-FALCLEIPIIWDLLYRREKILFGFGTCNGAELILNS 69
 QY 98 HVCACANIDIMTSMIHGYSNALHCHGSSPGSLQCIAGGVYASWFRVYDGAENOR 157
 DB 70 HCLLADA-----QMMOVHCHGSDSLQCLSTQVLAKEFLBEFVWGCVNKK 116
 QY 158 FIWREVVNTNMPREVMSSVFRGRLIYLLMTWDGHVS-----VVPAMSEGY 208
 DB 117 YLMREEVNSRDEINIVGSIEMRNHYIFRFSFVHOACMLAIORCISPGLGYF 176
 QY 209 SALHGLINNIIVLCCSYCADLSEIRVRCARTRRLMRAV-----IIAETTAMLY 262
 DB 177 KSTY---NWVLVKKCKSCSLQNYCALKSCAFWVRSIIDRVLRVEKIPVYLHRTS--- 230
 QY 263 SCRTERRRROQTRIRALLQHR 282
 DB 230 --KAERRRQALRQAMMGR 247
 RESULT 8
 E434_ADECC STANDARD: PRT: 265 AA.
 AC 065962;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE EARLY E4 31 KD PROTEIN.
 OS Canine adenovirus type 1 (strain CLL).
 OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Campbell J.B., Zhao Y.;
 RU Submitted (Aug-1996) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 248-265 FROM N.A.
 RX MEDLINE; 91272490.
 RA Draguliev B.P., Stira S., Abouhaider M.G., Campbell J.B.;
 RT "Sequence analysis of putative E3 and fiber genomic regions of two
 RL strains of canine adenovirus type 1."
 CC Virology 183:298-305(1991).
 CC -1- SIMILARITY: BELONGS TO THE ADENOVIRUS E4 30 TO 34 KD PROTEIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; U55001; AAB05452.1; -
 KW Early protein.
 SQ SEQUENCE 265 AA; 30822 MW; 92606E5E30134103 CRC64;
 Query Match 15.78; Score 250.5; DB 1; Length 265;
 Best Local Similarity 26.98; Pred. No. 3.1e-17;
 Matches 70; Conservative 43; Mismatches 96; Indels 51; Gaps 8;
 QY 51 CNTLMHNVSYVGRGP--CSVGFLLIOEWVPMVMYTRRELYI-----LRKM 97
 DB 11 CNAETSHVAVVAPAFPCNC-FALCLEIPIIWDLLYRREKILFGFGTCNGAELILNS 69
 QY 98 HVCACANIDIMTSMIHGYSNALHCHGSSPGSLQCIAGGVYASWFRVYDGAENOR 157
 DB 70 HCLLADA-----QMMOVHCHGSDSLQCLSTQVLAKEFLBEFVWGCVNKK 116
 QY 158 FIWREVVNTNMPREVMSSVFRGRLIYLLMTWDGHVS-----VVPAMSEGY 208
 DB 117 YLMREEVNSRDEINIVGSIEMRNHYIFRFSFVHOACMLAIORCISPGLGYF 176
 QY 209 SALHGLINNIIVLCCSYCADLSEIRVRCARTRRLMRAV-----IIAETTAMLY 262
 DB 177 KSTY---NWVLVKKCKSCSLQNYCALKSCAFWVRSIIDRVLRVEKIPVYLHRTS--- 230
 QY 263 SCRTERRRROQTRIRALLQHR 282
 DB 230 --KAERRRQALRQAMMGR 247
 RESULT 9
 E434_ADEM1 STANDARD: PRT: 289 AA.
 AC P23125;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE PROBABLE EARLY E4 33 KD PROTEIN (ORF A/B PROTEIN).
 OS Mouse adenovirus type 1 (MAV-1).
 OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 91082417.
 RA Ball A.O., Beard C.W., Villegas P., Spindler K.R.;
 RT "Early region 4 sequence and biological comparison of two isolates of
 RL mouse adenovirus type 1."
 Virology 180:257-265(1991).

CC -1- MISCELLANEOUS: THIS PROTEIN IS THE PRODUCT OF TWO DIFFERENT MRNAS
 CC SPLICED FROM THE SAME PRIMARY TRANSCRIPT.
 CC -1- SIMILARITY: BELONGS TO THE ADENOVIRUS E4 30 TO 34 KD PROTEIN
 CC FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC DR EMBL: M37187; AAA42514.1;
 CC DR PIR: A38519; 04ADM1.
 CC KM Early protein.
 CC SO SEQUENCE 289 AA; 33016 MW; F23D5C814DC63823 CRC64;

Query Match 11.7%; Score 186.5; DB 1; Length 289;
 Best Local Similarity 24.8%; Pred. No. 5.6e-11;
 Matches 68; Conservative 43; Mismatches 120; Indels 43; Gaps 11;

QY 52 NPLTHNYSYVRGLPCSGYFTLQEWV-----VPMQVLTREELVLRK 95
 DB 20 NPLTHNYSYVRGLPCSGYFTLQEWV-----VPMQVLTREELVLRK 77
 QY 96 CMHWCLCCANIDIMTSM--MIGYESMALHCHCSPGS-----LQCIAGGVYLAEMFR 146
 DB 78 L--DCLGFLPYCYQSFYSYVGTGDEVMISLHCHGROGSLQVPRRLQCLAAKARELYVQ 135
 QY 147 MYVDGAMNORIRIYREVYVNTNMPREVMSSVFMGRHLIYRLMYDGHVSYVPM-- 205
 DB 136 KELLGTRENEYPOYRHNANRYVNGLEYSYWC--GHFIYVRA--DGAFFALNKGIRA 192
 QY 205 SPFGSALHC--GILNNIVVLCSCYCA--DSEIRVCCARRRLMLRAVRIIAEETAM-- 261
 DB 193 RIGQGVLFPESLSCVYVIVCQACPPDQVDCMMLSTFLRMONLLGRGSSPL 252
 QY 261 -----LYSCRTERRCQFRLALLOHRIPLMHY 289
 DB 253 IPGFIDPRNRIERLRMLRREYRYTPYRLTY 286

RESULT 10
 UROM_BOVIN STANDARD: PRT: 643 AA.
 AC P48733;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-FEB-2000 (Rel. 39, Last annotation update)
 DE UROMODULIN PRECURSOR (TAMM-HORSFALL URINARY GLYCOPROTEIN) (THP).
 GN UMOD.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
 OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
 OC Bovinae; Bos.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-KIDNEY;
 RX MEDLINE: 95143938.
 RA Yu H., Papa F., Sukhatme V.P.;
 RT "Bovine and rodent tamm-horsfall protein (THP) genes: cloning,
 RT structural analysis, and promoter identification."
 RL Gene Expr. 4:63-75(1994).
 CC -1- FUNCTION: NOT KNOWN. MAY PLAY A ROLE IN REGULATING THE CIRCULATING
 CC ACTIVITY OF CYTOKINES AS IT BINDS TO IL-1, IL-2 AND TNF WITH HIGH
 CC AFFINITY (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR,
 CC THEN CLEAVED TO PRODUCE A SOLUBLE FORM WHICH IS SECRETED IN
 CC URINE (BY SIMILARITY).
 CC -1- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 ZP DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL: S75958; AAB33312.1;
 CC DR HSSP: P35555; IEMO.
 CC DR PRINTS: PR00023; ZPEBLUCIA.
 CC DR PROSITE: PS00682; ZP_DOMAIN; 1.
 CC DR PROSITE: PS00010; ASX_HYDROXYL; 2.
 CC DR PROSITE: PS00022; EGF_1; FALSE_NEG.
 CC DR PROSITE: PS01186; EGF_2; 3.
 CC DR PROSITE: PS01187; EGF_CA; 2.
 CC DR PFAM: PF00008; EGF; 3.
 CC DR PFAM: PF00100; zona_dellucida; 1.
 CC DR KW Glycoprotein; Signal; Membrane; GPI-anchor; EGF-like domain.
 CC FT SIGNAL 1 26
 CC FT CHAIN 27 643
 CC FT DOMAIN 30 66
 CC FT DOMAIN 67 109
 CC FT DOMAIN 110 151
 CC FT DOMAIN 336 587
 CC FT DISULFID 34 43
 CC FT DISULFID 37 52
 CC FT DISULFID 54 65
 CC FT DISULFID 71 85
 CC FT DISULFID 79 94
 CC FT DISULFID 96 108
 CC FT DISULFID 114 128
 CC FT DISULFID 122 137
 CC FT DISULFID 139 150
 CC FT CARBOHYD 40 40
 CC FT CARBOHYD 78 78
 CC FT CARBOHYD 134 134
 CC FT CARBOHYD 234 234
 CC FT CARBOHYD 277 277
 CC FT CARBOHYD 324 324
 CC FT CARBOHYD 398 398
 CC FT CARBOHYD 449 449
 CC FT CARBOHYD 515 515
 CC SQ SEQUENCE 643 AA; 69898 MW; 448984AB01DFA477 CRC64;

Query Match 5.7%; Score 91.5; DB 1; Length 643;
 Best Local Similarity 22.4%; Pred. No. 0.26;
 Matches 53; Conservative 19; Mismatches 86; Indels 79; Gaps 13;

QY 48 LPECNITLTHNYSYVRG-----LPCSVFTL-----IOEMVPMQVLTREEL 90
 DB 68 LDECAVIGAHNCSATKSCVNTLGSYTCVCEGLLSLSELCEDVDECAERG 119
 QY 91 VILKRCNHWCLCCANIDIMTSMHIGYESMALHCHCSPGS-----LQCIAGGVYLAEMFR 146
 DB 119 --LSRCHALATCINGEONYSYVAGYLGGRHCEC--SPSCGPGIDCVAREGDAL----- 171
 QY 147 MYVDGAMNORIRIYREVYVNTNMPREVMSSVFMGRHLIYRLMYDGHVSYVPM-- 204
 DB 171 VCVPEQCQVHRLIDYKSTEGSG--YICDVSLG-----WRYVGQAGVRLPET 218
 QY 205 SPFGSALHC-----GILNNIVVLCSCYCA--DSEIRVCCARRRLMLRAVRIIAEETAM-- 261
 DB 219 C--VPLHCNTAPRWMLNGTTPSSDEGVIRVACAHMSGDCLMDA---PIQYKACA 270

RESULT 11
 EAR_EBV STANDARD: PRT: 191 AA.
 ID EAR_EBV
 AC P03182;
 DT 21-JUL-1986 (Rel. 01, Created)

```

Db      119 QSPFYVVDLSVAGMFLASEJLDCW-HIQQ-----GC-----NSTLIEDNIPGS 161
Qt      156 QRFIWRREVYNNMPREVMSSVFMRGRH 185
          ||| | : | : |||
Db      162 RRFSTWLEFAGILTLSLVLYCSYLFI-SRGRH 191
          ||| | : | : |||

RESULT 12
OPSD_XENLA ID OPSD_XENLA STANDARD: PRT; 354 AA.
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DE 15-JUL-1999 (Rel. 38, Last annotation update)
DE RHODOPSIN.
GN RHO.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;
OC Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae; Xenopodinae;
Xenopus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 93287804.
RA Saha M.S., Gralinger R.M.;
RT "Early opsin expression in xenopus embryos precedes photoreceptor
RT differentiation.";
RL Brain Res. Mol. Brain Res. 17:307-318(1993).
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 96216396.
RA Banti S., Scalzetti L.C., Moody S.A., Knox B.E.;
RT "Characterization of the xenopus rhodopsin gene.";
RL J. Biol. Chem. 271:3179-3186(1996).
-1- FUNCTION: VISUAL PIGMENTS ARE THE LIGHT-ABSORBING MOLECULES THAT
MEDIANE VISION. THEY CONSIST OF AN APOPROTEIN, OP SIN, COVALENTLY
LINKED TO CIS-RETINAL.
-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
-1- TISSUE SPECIFICITY: ROD SHAPED PHOTORECEPTOR CELLS WHICH MEDIATES
VISION IN DIM LIGHT.
-1- PTM: SOME OR ALL OF THE CARBOXYL-TERMINAL SER OR THR RESIDUES MAY
BE PHOSPHORYLATED.
-1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
OSIN SUBFAMILY.
-----+-----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sdb.ch/announce/
or send an email to license@isb-sib.ch).
-----+-----
DR EMBL; S62292; CAB32105.1; -
DR EMBL; I04692; AAB59590.1; -
DR EMBL; I07770; AAC42232.1; -
DR EMBL; U23808; AAC59901.1; -
DR PIR; S31473; S31473.
DR GCRDB; GCR_0487; -
DR GCRDB; GCR_0497; -
DR GCRDB; GCR_0972; -
DR PRINTS; PR00238; OPSIN.
DR PRINTS; PR00579; RHODOPSIN.
DR PROSITE; PS00237; G_PROTEIN_RECEPTOR; 1.
DR PROSITE; PS00238; OPSIN; 1.
DR Pfam; PF00001; 7tm_1; 1.
KW Photoreceptor; Retinal protein; Transmembrane; Glycoprotein; Vision;
KW Phosphorylation; Lipoprotein; Palmitate; G-protein coupled receptor.
FT CARBOHYD 2 BY SIMILARITY.
FT DOMAIN 15 BY SIMILARITY.
FT TRANSMEM 37 EXTRACELLULAR.
FT DOMAIN 61 1 (POTENTIAL).
FT CYTOPLASMIC.
FT TRANSMEM 74 2 (POTENTIAL).

```

FT DOMAIN 99 113 EXTRACELLULAR.
 FT TRANSMEM 114 133 3 (POTENTIAL).
 FT DOMAIN 134 152 CYTOPLASMIC.
 FT TRANSMEM 153 176 4 (POTENTIAL).
 FT DOMAIN 177 202 5 (POTENTIAL).
 FT TRANSMEM 203 230 5 (POTENTIAL).
 FT DOMAIN 231 252 6 (POTENTIAL).
 FT TRANSMEM 253 276 7 (POTENTIAL).
 FT DOMAIN 277 284 7 (POTENTIAL).
 FT TRANSMEM 285 309 7 (POTENTIAL).
 FT DOMAIN 310 334 7 (POTENTIAL).
 FT BLINDING 322 322 RETINAL CHROMOPHORE.
 FT LIPID 323 323 PALMITATE (BY SIMILARITY).
 FT LIPID 323 323 PALMITATE (BY SIMILARITY).
 FT DISULFID 110 187 BY SIMILARITY.
 FT CONFLICT 107 107 P -> Q (IN REF. 2).
 FT CONFLICT 137 137 I -> M (IN REF. 2).
 FT CONFLICT 241 241 L -> A (IN REF. 2).
 SQ SEQUENCE 354 AA; 39786 MW; CD18F49EC63D8FE CRC64;

Query Match 5.18; Score 81.5; DB 1; Length 354;
 Best Local Similarity 20.0%; Pred. No. 1.2;
 Matches 61; Conservative 33; Mismatches 94; Indels 117; Gaps 13;

QY 5 GVPFG-MTLRPTSRSLSRPYSRDLPPETETRAITLEDHPLPCCNTLTHANV----- 60
 DB 51 GLPIFMFTLFTIQHKRLKPLNTYLL-----NLVFNHFVYLCGFVTMTSMHGY 102
 QY 60 -----SYVGLGPCSVFTLIQENVYPMADVLTRELYILRKCHVCLCCANIDIMTSM 112
 DB 103 FIFGPTGCTIEF-----FATLGEVALMSLVLAVERIYV-----VCKPMANRFGENH 152
 QY 113 MHGYE-SWALHCHSSPSGLQIAGGOVLASWFRMYVDG-----AMENQR 157
 DB 153 AIMGAFMTIMLSCAP-----PLFGMSRYIPEGMCSCGVYDYLKREVNES 202
 QY 158 FTWYREVYNYNPKREVMSSVFMGRHL----- 187
 DB 203 FVLYMEIVHFTPLIYIF-----FCYGRLLCTYKEAQQOESLTOKAEKVTIRVAVIN 258
 QY 187 -----ITLRMYDGHVS-----VYPAMSFYSALHCG-----ILNN-----I 219
 DB 259 VFPLIGWYVAIVAFITFHGSGNFGPVMTVPAPFAKSSALINPVYIYLKQFRNCLI 318
 QY 220 VVLCC 224
 DB 319 TILCC 323

RESULT 13
 OPSD_LIMBE STANDARD: PRT; 289 AA.
 AC 042427;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE RHODOPSIN (FRAGMENT).
 GN RHO.
 OS Limnocottus bergianus.
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Actinopterygii;
 OC Neopterygii; Teleostei; Euteleostei; Acanthopterygii; Percomorphia;
 OC Scorpaeniformes; Cottoidae; Abyssocottidae; Limnocottus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE; 98086781.
 RA Hunt D.M., Fitzgibbon J., Slobodanyuk S.J., Bowmaker J.K.,
 RA Dzial K.S.;
 RA "Molecular evolution of the cottoid fish endemic to Lake Baikal
 RA deduced from nuclear DNA evidence".
 RL Mol. Phylogenet. Evol. 8:415-422(1997).
 CC -1- FUNCTION: VISUAL PIGMENTS ARE THE LIGHT-ABSORBING MOLECULES THAT
 CC MEDIATE VISION. THEY CONSIST OF AN APOPROTEIN, OPSIN, COVALENTLY

CC LINKED TO CIS-RETINAL.
 CC -1- SURCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: ROD SHAPED PHOTORECEPTOR CELLS WHICH MEDIATES
 CC VISION IN DIM LIGHT.
 CC -1- PTM: SOME OR ALL OF THE CARBOXYL-TERMINAL SER OR THR RESIDUES MAY
 CC BE PHOSPHORYLATED (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CC OPSIN SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; U97270; AAB61724.1; -.
 DR GCRDB; GCR_1264; -.
 DR PROSITE; PS00237; G-PROTEIN-RECEPTOR; 1.
 DR PROSITE; PS00238; OPSIN; 1.
 DR PFM; PFM0001; 7tm.1; 1.
 KW Photoreceptor; Retinal protein; Transmembrane; Glycoprotein; Vision;
 KW Phosphorylation; Lipoprotein; Palmitate; G-protein coupled receptor.
 FT NON_TER 1 1
 FT DOMAIN <1 7
 FT TRANSMEM 8 32 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 33 44 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 45 69 2 (POTENTIAL).
 FT DOMAIN 70 84 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 85 104 3 (POTENTIAL).
 FT TRANSMEM 105 123 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 124 147 4 (POTENTIAL).
 FT TRANSMEM 148 173 5 (POTENTIAL).
 FT TRANSMEM 174 201 5 (POTENTIAL).
 FT DOMAIN 202 223 6 (POTENTIAL).
 FT TRANSMEM 224 247 6 (POTENTIAL).
 FT TRANSMEM 248 255 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 256 280 7 (POTENTIAL).
 FT DOMAIN 281 >289 CYTOPLASMIC (POTENTIAL).
 FT DISULFID 81 158 BY SIMILARITY.
 FT BLINDING 267 267 RETINAL CHROMOPHORE (BY SIMILARITY).
 FT CARBOHYD 171 171
 FT NON_TER 289 289
 SQ SEQUENCE 289 AA; 32445 MW; 1B3812014CB80A2 CRC64;

Query Match 5.0%; Score 79; DB 1; Length 289;
 Best Local Similarity 24.0%; Pred. No. 1.7;
 Matches 36; Conservative 22; Mismatches 54; Indels 38; Gaps 8;

QY 55 TMHNSYVRLGIPSV-GF--TLIQENVYPMADVLTRELYILRKCHVCLCCANIDIMTS 111
 DB 69 SMHGYSVLGRGLIEGFATLGE--IALMSLV-----VALERWYVCKPISNREFETD 122
 QY 112 MHGYE-SWALHCHSSPSGLQIAGGOVLASWFRMYVDG-----AMFNQ 156
 DB 123 HAINGLAFSWMALACAP-----PLVGRSRIIPBGMCSGVDYTRAEGFNN 172
 QY 157 REIYREVYNYNPKREVMSSVFMGRHL 186
 DB 173 SFVIMFTVHLILPLSVF-----FCYGRLL 198
 RESULT 14
 OPSD_POMMI STANDARD: PRT; 352 AA.
 ID OPSD_POMMI
 AC P35403;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE RHODOPSIN.
 GN RHO.

OS Pomatoschistus minutus (Sand goby).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
 CC Neopterygii; Teleostei; Euteleostei; Acanthopterygii; Percomorpha;
 CC Perciformes; Gobioidae; Gobiidae; Pomatoschistus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=RETINA;
 RX MEDLINE; 92396703.
 RA Archer S.N., Lythgoe J.N., Hall L.;
 RT "Rod opsin cDNA sequence from the sand goby (Pomatoschistus minutus)
 RL compared with those of other vertebrates";
 RL Proc. R. Soc. Lond. B. Biol. Sci. 246:19-25(1992).
 CC -1- FUNCTION: VISUAL PIGMENTS ARE THE LIGHT-ABSORBING MOLECULES THAT
 CC MEDIANE VISION. THEY CONSIST OF AN APOPROTEIN, OPSIN, COVALENTLY
 CC LINKED TO CIS-RETINAL.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: ROD SHAPED PHOTORECEPTOR CELLS WHICH MEDIATES
 CC VISION IN DIM LIGHT.
 CC -1- PTM: SOME OR ALL OF THE CARBOXYL-TERMINAL SER OR THR RESIDUES MAY
 CC BE PHOSPHORYLATED.
 CC -1- MISCELLANEOUS: THIS RHODOPSIN HAS AN ABSORPTION MAXIMA AT 501 NM.
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CC OPSIN SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X62405; CA44275.1; .
 DR PIR; S40688; S40688.
 DR GCRDB; GCR_0503; .
 DR PRINTS; PR00237; GPCRHHODPSN.
 DR PRINTS; PR00238; OPSIN.
 DR PRINTS; PR00579; RHODOPSIN.
 DR PROSITE; PS00237; G-PROTEIN_RECEPTOR; 1.
 DR PROSITE; PS00238; OPSIN; 1.
 DR PFAM; PF00001; 7tm1; 1.
 KW Photoreceptor; Retinal protein; Transmembrane; Glycoprotein; Vision;
 RV phosphorylation; Lipoprotein; Palmitate; G-protein coupled receptor.
 FT DOMAIN 1
 FT TRANSSEM 37 61
 FT TRANSSEM 62 73
 FT TRANSSEM 74 98
 FT TRANSSEM 99 113
 FT TRANSSEM 114 133
 FT TRANSSEM 134 152
 FT TRANSSEM 153 176
 FT TRANSSEM 177 202
 FT TRANSSEM 203 230
 FT TRANSSEM 231 252
 FT TRANSSEM 253 276
 FT TRANSSEM 277 284
 FT TRANSSEM 285 309
 FT TRANSSEM 310 352
 FT DOMAIN 2
 FT CARBOHYD 15 15
 FT BINDING 296 296
 FT LIPID 322 322
 FT LIPID 323 323
 FT DISULFID 110 187
 SQ SSOURCE 352 AA; 39409 MW; 88BCCE379A9ED656 CRC64;

Query Match 5.0%; Score 79; DB 1; Length 352;
 Best Local Similarity 19.9%; Pred. No. 2.1;
 Matches 51; Conservative 31; Mismatches 76; Indels 98; Gaps 13;

55 THANVSFRLPCSV-GF-TLIGENVYPMQVTLRELYILRKCMHYCLCANIDIMTS 111

Db 98 SMHGFEVLGRGCVNVEGFALLIGE-IALMSLVLAVERMV-----VCKPISNRFETEN 151
 QY 112 MMRIGE-SWALHCHSSPSPSLGCIINGGVLAIFRMVVDG-----AMFNQ 156
 Db 152 HAIGVAFSWIMATCAVP-----PLVGMSKTYPEGMQSCGVYTYRARGFENNE 201
 QY 157 RFIWREVVYNNMPKEVMWSVFMRGRHLYLR----- 191
 Db 202 SFVLYMEIVHFLAPLYIVF-----FCYGRLLCAVKKAAQAQSEETTORAREEVYRMTYM 257
 QY 191 -----LWY-----DGHGVSVPAMSGYSALHCGILINNIIVLCCS--- 226
 Db 258 VIGFLISWLPYASVAVYITHTGTEFGRPLFMIPAFKSSALY-----NPMIYICMKNOF 313
 QY 226 -YCADISEIRVRCAR 240
 Db 314 RHCM-----ITTLCCGR 325

RESULT 15
 ID OPSD_ZEUFA STANDARD; PRT; 354 AA.
 AC 042604;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1998 (Rel. 38, Last annotation update)
 DE RHODOPSIN.
 GN RHO.
 OS Zeus faber ('John Dory').
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
 CC Neopterygii; Teleostei; Euteleostei; Acanthopterygii; Percomorpha;
 CC Perciformes; Zeidae; Zeus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=LIVER;
 RA Archer S.N., Hirano J.;
 RL Submitted (Aug-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: VISUAL PIGMENTS ARE THE LIGHT-ABSORBING MOLECULES THAT
 CC MEDIANE VISION. THEY CONSIST OF AN APOPROTEIN, OPSIN, COVALENTLY
 CC LINKED TO CIS-RETINAL.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: ROD SHAPED PHOTORECEPTOR CELLS WHICH MEDIATES
 CC VISION IN DIM LIGHT.
 CC -1- PTM: SOME OR ALL OF THE CARBOXYL-TERMINAL SER OR THR RESIDUES MAY
 CC BE PHOSPHORYLATED.
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CC OPSIN SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Y14484; CAA74832.1; .
 DR GCRDB; GCR_2445; .
 DR PRINTS; PR00237; GPCRHHODPSN.
 DR PRINTS; PR00238; OPSIN.
 DR PRINTS; PR00579; RHODOPSIN.
 DR PROSITE; PS00237; G-PROTEIN_RECEPTOR; 1.
 DR PROSITE; PS00238; OPSIN; 1.
 DR PFAM; PF00001; 7tm1; 1.
 KW Photoreceptor; Retinal protein; Transmembrane; Glycoprotein; Vision;
 RV phosphorylation; Lipoprotein; Palmitate; G-protein coupled receptor.
 FT DOMAIN 1
 FT TRANSSEM 37 61
 FT TRANSSEM 62 73
 FT TRANSSEM 74 98
 FT TRANSSEM 99 113
 FT TRANSSEM 114 133
 FT TRANSSEM 134 152
 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 153 176 4 (POTENTIAL).
 FT DOMAIN 177 202 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 203 230 5 (POTENTIAL).
 FT DOMAIN 231 252 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 253 276 6 (POTENTIAL).
 FT DOMAIN 277 284 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 285 309 7 (POTENTIAL).
 FT DOMAIN 310 354 CYTOPLASMIC (POTENTIAL).
 FT CARBOHYD 2 2 BY SIMILARITY.
 FT CARBOHYD 15 15 BY SIMILARITY.
 FT DISULFID 110 187 BY SIMILARITY.
 FT BINDING 296 296 RETINAL CHROMOPHORE (BY SIMILARITY).
 FT LIPID 322 322 PALMITATE (BY SIMILARITY).
 FT LIPID 323 323 PALMITATE (BY SIMILARITY).
 SQ SEQUENCE 354 AA; 39662 MW; A97D38F8E86DB3 CRC64;

Query Match 5.0%; Score 79; DB 1; Length 354;
 Best Local Similarity 18.7%; Pred. No. 2.1;
 Matches 53; Conservative 35; Mismatches 87; Indels 108; Gaps 14;

OY 55 TWNVSYVGLPCSV-GE--TLIOEVVPMQVLTREELVLRKCMHYCLCCANIDMTS 111
 DB 98 SMHGFVILGRGNCNLEGFATLGE-IALWSIVLAVERNVY---VCKPISNFRGEN 151
 OY 112 KMIHGE-SMALHCHCSSPGSLQCIAGQVLAWEFRMVVDGAMF-----NQ 156
 DB 152 HAVMGVSTWLAACASVP-----PLFGMSRYIPEGMQCSGIDYTRAPGVNNE 201
 OY 157 REIWRREVNNMPKREVMMSVFMRGRHLI----- 188
 DB 202 SPVITMFVCHFSIPLTIIP---FCYGRLLCAVKDAAAOESETTORAREVSRMYIM 257
 OY 188 ---YLRWTD-----GHWGSVVPAMSFYSALHCGILNNIVLCCS--- 226
 DB 258 VIGFLICMLPYASVAMFIFTHOGSEFGPVEMTIPA---FFAKSSAIYNPMIYICMKNQF 313
 OY 226 -YCADLSEIRVRCARTRRLMLRAVRIAEETLMLYSCRT 267
 DB 314 RHCM---ITTLCCGKNP-----FEEEGASTASKTE 342

Search completed: June 23, 2000, 10:14:48
 Job time: 669 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2000, 09:51:57 ; Search time 61.44 "Seconds
(without alignments)
331.775 Million cell updates/sec

Title: US-09-214-478-2

Perfect score: 1594
Sequence: 1 MTSGVPPGKMLRPTSRSL...RALLQHRPILMHVDSTPM 294

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 225878 seqs, 69334122 residues

Total number of hits satisfying chosen parameters: 225878

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : SPTREMBL.12.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_ricent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	223.5	14.0	219	071106	071106 bovine aden
2	180	11.3	163	084211	084211 porcine ade
3	180	11.3	163	09YTR0	09YTR0 porcine ade
4	163.5	10.3	268	071108	071108 bovine aden
5	108	6.8	220	P89027	P89027 ovine adeno
6	90	5.6	1293	09Y748	09Y748 emericella
7	87	5.5	494	044515	044515 caenorhabd1
8	85.5	5.4	211	067148	067148 aquifex neo
9	85.5	5.4	954	097180	097180 drosophila
10	84.5	5.3	431	023017	023017 arabidopsis
11	83	5.2	354	10	023666 arabidopsis
12	82.5	5.2	181	024416	024416 fragaria an
13	82.5	5.2	405	021884	021884 plasmodium
14	82	5.1	1603	097285	097285 sparus aua
15	80.5	5.0	353	022130	022130 caenorhabd1
16	80	5.0	456	0218735	0218735 canis fam11
17	79.5	5.0	1259	09YH05	09YH05 diplopus an
18	79.5	5.0	353	09YH03	09YH03 sarpa salpa
19	79	5.0	353	13	09YH03
20	79	5.0	353	13	09YH03

21	78	4.9	353	13	09YH04	09YH04 diplopus vu
22	77.5	4.9	459	8	092258	092258 canis fam11
23	77.5	4.9	530	12	09YH03	09YH03 lymantria d
24	77.5	4.9	642	12	0928V7	0928V7 streptomyce
25	77.5	4.9	3008	12	039929	039929 hepatitis c
26	77	4.8	227	10	023668	023668 arabidopsis
27	77	4.8	353	13	09YH00	09YH00 lithognathu
28	76	4.8	505	2	059435	059435 escherichia
29	76	4.8	932	10	082748	082748 arabidopsis
30	75.5	4.7	373	5	022188	022188 caenorhabd1
31	75.5	4.7	407	12	040917	040917 kaposi's sa
32	74.5	4.7	303	5	017080	017080 caenorhabd1
33	74.5	4.7	321	3	008959	008959 saccharomyc
34	74.5	4.7	407	12	P88909	P88909 kaposi's sa
35	74.5	4.7	895	5	045631	045631 caenorhabd1
36	74	4.6	353	13	09Y628	09Y628 chelon labr
37	74	4.6	353	13	09Y627	09Y627 liza salien
38	74	4.6	3072	12	092645	092645 clover yell
39	73.5	4.6	295	12	011425	011425 avian adeno
40	73.5	4.6	315	4	050403	050403 homo sapien
41	73.5	4.6	346	8	021195	021195 megachasma
42	73.5	4.6	348	11	09YH01	09YH01 cavia porce
43	73	4.6	353	13	09Y629	09Y629 mugil cepha
44	73	4.6	353	13	09Y626	09Y626 liza aurata
45	73	4.6	367	5	P91382	P91382 caenorhabd1

ALIGNMENTS

RESULT 1
ID 071106 PRELIMINARY: PRT: 219 AA.

AC 071106;
DT 01-AUG-1998 (TREMBLrel. 07, last sequence update)
DT 01-AUG-1998 (TREMBLrel. 07, last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, last annotation update)
DE ADENOVIRUS 3 COMPLETE GENOME.
OS Bovine adenovirus type 3 (Mastadenovirus bovis).
OC Bovine adenovirus type 3 (Mastadenovirus).
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MBR-1;
RA LEE J.B., BAXI M.K., IDAMAKANTI N., REDDY P.S., ZAKHARTCHOUK A.N.,
RA PINE C., BABIUK L.A., TIKOO S.K.;
RL Virus Genes 0:0-0(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-MBR-1;
RA BAXI M.K., REDDY P.S., ZAKHARTCHOUK A.N., IDAMAKANTI N., PINE C.,
RA BABIUK L.A., TIKOO S.K.;
RL Virus Genes 0:0-0(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-MBR-1;
RA MEDLINE: 98105785.
RA REDDY P.S., IDAMAKANTI N., ZAKHARTCHOUK A.N., BAXI M.K., LEE J.B.,
RA PINE C., BABIUK L.A., TIKOO S.K.;
RL "Nucleotide sequence, genome organization, and transcription map of
RT bovine adenovirus type 3".
RL J. Virol. 72:1394-1402(1998).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-MBR-1;
RA REDDY P.S., IDAMAKANTI N., ZAKHARTCHOUK A.N., BAXI M.K., LEE J.B.,
RA PINE C., BABIUK L.A., TIKOO S.K.;
RL Submitted (Oct-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF030154; AAD09737.1;
SQ SEQUENCE 219 AA; 26028 MW; 092D5CCD CRC32;

Query Match 14.0%; Score 223.5; DB 12; Length 219;
Best Local Similarity 25.3%; Pred. No. 7.7e-16;

Matches 56; Conservative 42; Mismatches 106; Indels 17; Gaps 5;

QY 77 WVPMDVLTREELVIRKCHV--CLC--CANIDI--TSMHIGESWALHCHCSP 129
 Db 4 WIPKXVIMPEP-----AHLSCLCDYCLFVIGKRKRSYVGEELVYHCHCPK 56
 QY 130 GSLOCAGGCVLASWFMVVDGAMFNORFTYREVYNNKFKEMSSVFMGRHLTYL 189
 Db 57 RSLQCRSAAYVFGVLCVQVHGAQFDRVYVYFRELSTKSLSVFLAYGSVYNNVHLTYF 116
 QY 190 RLVWDGHSVVPAMSGYALHCGILNNIVYLCSTCADLSEIRVRCARRTRMLRA 249
 Db 117 NSTFMIEGSTRCTPGECLWKGNSCIYLLICRSCNPLEMSVNCCLKCKOKIRM 176
 QY 250 VRIAEETAMLYSCRTERRROOFIRALIDHNPILMHEDY 290
 Db 177 LNLAPKRVKHL---GLARLRDRFLRLARFREPVLFDNRD 214

RESULT 2
 084211 PRELIMINARY; PRT; 163 AA.
 AC 084211;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DT 01-NOV-1996 (TREMblrel. 01, Last annotation update)
 DE MASTADENOVIRUS 19.8 KDA POLYPEPTIDE (ORF 8).
 OS Porcine adenovirus 3.
 OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-PAV-3;
 RA REDDY S., IDAMAKANTI N., DERBYSHIRE J.B., NAGY E.;
 DR EMBL: L43363; CAB24462.1; -
 KW Repeat.
 SQ SEQUENCE 163 AA; 19786 MW; DAE89A1D CRC32;

Query Match
 Best Local Similarity 11.3%; Score 180; DB 12; Length 163;
 Matches 45; Conservative 20; Mismatches 49; Indels 28; Gaps 4;

QY 141 LASFRMYVDGAMFNORFTYREVYNNKFKEMSSVFMGRHLTYLWTD----- 195
 Db 1 MRSIMREIIRCEYRIRTWFRBAVNLPSVRYVGSVFMGRHLYIKIMYDDCKLL 60
 QY 195 --GHVGSVVPAMSGYALHCGILNNIVYLCSTCADLSEIRVRCARRTRMLRAV 251
 Db 61 RMFMGEVI-----YCGDYANYIVLCRCHELSEVPARCAQRLHMKLAAE 110
 QY 252 IIAEETAMLYSCRTERRROOF 273
 Db 111 ALGAORRG-----RKQOF 123

RESULT 3
 09YTR0 PRELIMINARY; PRT; 163 AA.
 AC 09YTR0;
 DT 01-MAY-1999 (TREMblrel. 10, Created)
 DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)
 DT 01-MAY-1999 (TREMblrel. 10, Last annotation update)
 DE 19.8 KDA POLYPEPTIDE.
 OS Porcine adenovirus 3.
 OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-6618;
 RA REDDY P.S., NAGY E., DERBYSHIRE J.B.;
 RT "Sequence analysis of putative pVIIT, E3 and fibre regions of porcine adenovirus type 3.";

RL Virus Res. 36:97-106(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-6618;
 RX MEDLINE; 95407103.
 RA REDDY P.S., TUDOLY T., DENNIS J.R., DERBYSHIRE J.B., NAGY E.;
 RT "Comparison of the inverted terminal repetition sequences from five porcine adenovirus serotypes.";
 RL Virology 212:237-239(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-6618;
 RX MEDLINE; 98014367.
 RA REDDY P.S., IDAMAKANTI N., DERBYSHIRE J.B., NAGY E.;
 RT "Characterization of the early region 4 of porcine adenovirus type 3.";
 RL Virus Genes 15:87-90(1997).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-6618;
 RX MEDLINE; 99058191.
 RA REDDY P.S., IDAMAKANTI N., SONG J.Y., LEE J.B., HYUN B.H., PARK J.H.,
 RA CHA S.H., BAE Y.T., TIKOO S.K., BABIUK L.A.;
 RT "Nucleotide sequence and transcription map of porcine adenovirus type 3.";
 RL Virology 251:414-426(1998).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN-6618;
 RA REDDY P.S.;
 RT "Sequence and transcription map analysis of early region-1 of porcine adenovirus type 3.";
 RL Virus Res. 0:0-0(1999).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN-6618;
 RA REDDY P.S.;
 RT Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF083132; AAC39443.1; -
 SQ SEQUENCE 163 AA; 19786 MW; DAE89A1D CRC32;

Query Match
 Best Local Similarity 11.3%; Score 180; DB 12; Length 163;
 Matches 45; Conservative 20; Mismatches 49; Indels 28; Gaps 4;

QY 141 LASFRMYVDGAMFNORFTYREVYNNKFKEMSSVFMGRHLTYLWTD----- 195
 Db 1 MRSIMREIIRCEYRIRTWFRBAVNLPSVRYVGSVFMGRHLYIKIMYDDCKLL 60
 QY 195 --GHVGSVVPAMSGYALHCGILNNIVYLCSTCADLSEIRVRCARRTRMLRAV 251
 Db 61 RMFMGEVI-----YCGDYANYIVLCRCHELSEVPARCAQRLHMKLAAE 110
 QY 252 IIAEETAMLYSCRTERRROOF 273
 Db 111 ALGAORRG-----RKQOF 123

RESULT 4
 071108 PRELIMINARY; PRT; 268 AA.
 AC 071108;
 DT 01-AUG-1998 (TREMblrel. 07, Created)
 DT 01-AUG-1998 (TREMblrel. 07, Last sequence update)
 DT 01-NOV-1998 (TREMblrel. 08, Last annotation update)
 DE ADENOVIRUS 3 COMPLETE GENOME.
 OS Bovine adenovirus type 3 (Mastadenovirus bos3).
 OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-WB-1;
 RA LEE J.B., BAXI M.K., IDAMAKANTI N., REDDY P.S., ZAKHARCHOUK A.N.,

RA PYNE C., BABIUK L.A., TIKOO S.K.;
 RL Virus Genes 0:0-0(1998).
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-WB-1;
 RA BAXI M.K., REDDY P.S., ZAKHARCHOUK A.N., IDAMAKANTI N., PYNE C.,
 BABIUK L.A., TIKOO S.K.;
 RL Virus Genes 0:0-0(1998).
 [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-WB-1;
 RA MEDLINE; 98105785.
 RA REDDY P.S., IDAMAKANTI N., ZAKHARCHOUK A.N., BAXI M.K., LEE J.B.,
 PYNE C., BABIUK L.A., TIKOO S.K.;
 RT "Nucleotide sequence, genome organization, and transcription map of
 RT bovine adenovirus type 3";
 RL J. Virol. 72:1394-1402(1998).
 [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-WB-1;
 RA REDDY P.S., IDAMAKANTI N., ZAKHARCHOUK A.N., BAXI M.K., LEE J.B.,
 PYNE C., BABIUK L.A., TIKOO S.K.;
 RT Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF030154; AAD09739.1;
 SQ SEQUENCE 268 AA; 31016 MW; BEAA16AF CRC32;
 Query Match 10.3%; Score 163.5; DB 12; Length 268;
 Best Local Similarity 22.0%; Pred. No. 2e-09; Mismatches 92; Indels 17; Gaps 6;
 Matches 44; Conservative 47;
 OY 100 CACGANNIDMTSMHIGYSNALHCHGSSPG--SLQCIAGGQVLAWEFVVDGAMFNOR 157
 DB 59 CDOCTIRGCDSEFFISGDAITWRYC-CSKADQOSLOGLCRATILMDLFTMCVAGAFDM 117
 OY 158 FIVREVYVNYNMEKVEVMSVFMGRHLIYLTMDTHVGS-----VVPMSFGYSAL 211
 DB 118 FPFRELSNLSRTYIYVGVYVNRVHLLY---FTNVGASNOLKSLPFANGNYM 173
 OY 212 HCGILNNIYVLCSCYCADSEIRVRCARTRMLRAVRIIAETMTALYSCRTERRRO 271
 DB 174 GSGLDKCIIPCYDAELSEVASRCLKVKEMIMCYITCLPSDRH---YOCGLEPHND 230
 OY 272 QFIRALLQHRPIIMHDYDS 291
 DB 231 KLIQRYAA-GEFVLRKVFDT 249
 RESULT 5
 P89027 PRELIMINARY; PRT; 220 AA.
 AC P89027;
 DT 01-MAY-1997 (T-EMBLrel. 03, Created)
 DT 01-NOV-1999 (T-EMBLrel. 12, Last sequence update)
 DT 01-NOV-1999 (T-EMBLrel. 12, Last annotation update)
 DE E4.2 PROTEIN (25.6 KD PROTEIN).
 OS Ovine adenovirus.
 OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Adenovirus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-OAV287;
 RA MEDLINE; 96240641.
 RA YRATI S., BROOKES D.E., STRIKE P., KHATRI A., BOYLE D.B., BOTH G.W.;
 RT "Unique genome arrangement of an ovine adenovirus: identification of
 RT new proteins and proteinase cleavage sites";
 RL Virology 220:186-199(1996).
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-OAV287;
 RA MEDLINE; 97271311.
 RA XU Z.Z., HYATT A., BOYLE D.B., BOTH G.W.;
 RT "Construction of ovine adenovirus recombinants by gene insertion or
 RT deletion of related terminal region sequences.";

RL Virology 230:62-71(1997).
 [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-OAV287;
 RX MEDLINE; 98277079.
 RA KHATRI A., BOTH G.W.;
 RT "Identification of transcripts and promoter regions of ovine
 RT adenovirus OAV287";
 RL Virology 245:128-141(1998).
 [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-OAV287;
 RA BOTH G.W.;
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U40839; AAD45954.1;
 SQ SEQUENCE 220 AA; 25401 MW; DCAF782 CRC32;
 Query Match 6.8%; Score 108; DB 12; Length 220;
 Best Local Similarity 27.7%; Pred. No. 0.0011;
 Matches 43; Conservative 22; Mismatches 50; Indels 40; Gaps 10;
 OY 123 HCHGSSPGSLQCIAGGQVLAWEFVVDGAMFNORFIWYREV-VYNNPKVEVMSVFM 180
 DB 78 HCHGSSPGSLQCIAGGQVLAWEFVVDGAMFNORFIWYREV-VYNNPKVEVMSVFM 135
 OY 181 MGRHLIYLT-----RLMTDGHVGSVVPMSFGYSALHCGIINNYVLCSCYCADL 230
 DB 136 MGRHLIYLT-----RLMTDGHVGSVVPMSFGYSALHCGIINNYVLCSCYCADL 184
 OY 231 SEIRVRCARTRMLRAVRIIAETMTALYSC 264
 DB 185 LQ---QCISREKFNFLR-----MLYCC 205
 RESULT 6
 Q9Y748 PRELIMINARY; PRT; 1293 AA.
 AC Q9Y748;
 DT 01-NOV-1999 (T-EMBLrel. 12, Created)
 DT 01-NOV-1999 (T-EMBLrel. 12, Last sequence update)
 DT 01-NOV-1999 (T-EMBLrel. 12, Last annotation update)
 DE ABC TRANSPORTER PROTEIN ATRC.
 GN ATRC.
 OS Emericella nidulans (Aspergillus nidulans).
 OC Eukaryota; Fungi; Ascomycota; Euscomycetes; Plecomycetes;
 CC Eurotiales; Trichocomaceae; Emericella.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-FGSC A4;
 RX MEDLINE; 99156773.
 RA ANGERMAYR K., PARSON W., STOFFLER G., HAAS H.;
 RT "Expression of atrc - encoding a novel member of the ATP binding
 RT cassette transporter family in Aspergillus nidulans - is sensitive to
 RT cycloheximide";
 RL Biochim. Biophys. Acta 1453:304-310(1999).
 DR EMBL; AF082072; AAD25925.1;
 DR PROSITE; PS00211; ABC_TRANSPORTER; 2.
 KW ATP-binding; Transport.
 SQ SEQUENCE 1293 AA; 141984 MW; 2D8E0446 CRC32;
 Query Match 5.6%; Score 90; DB 3; Length 1293;
 Best Local Similarity 21.4%; Pred. No. 0.65;
 Matches 50; Conservative 36; Mismatches 94; Indels 54; Gaps 10;
 OY 90 LVIRKCMVCLCCAN-----IDMTSMHIGYSNALHCHGSSPGSLQCI 136
 DB 717 LVIRKCMVCLCCAN-----IDMTSMHIGYSNALHCHGSSPGSLQCI 775
 OY 137 G-----GQVLAWEFVVDGAMFNORFIWYREV-VYNNPKVEVMSVFM 179
 DB 776 GIGFEFTVAASFLSGYRIRYFAAMLNDVSFEEDDSAGVMTGSLTDIPRIEDLISL 835

QY 180 FMGRHLLIYRLMTDYGIVVPMSPGYSALHGLINNIYVLCSCADLSLRV-CC 238
 DB 836 CLGFIIVVAV-----JASCIILAVGWRALVAIFGCPPL---FLAYVRRLEITC 887
 QY 239 ARFRRLMLR-----AVRIIA-----EETAMJSCRTERRROOFTALL 278
 DB 888 QERTRLYESARFATEAISAFITVASILTEEKVIOMYDRLSHTSPKFRITL 941

RESULT 7

AC 044515 PRELIMINARY; PRT; 494 AA.
 ID 044515

DT 01-JUN-1998 (Tremblrel. 06, Created)
 DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
 DT 01-NOV-1998 (Tremblrel. 08, Last annotation update)
 DE F56A11.4 PROTEIN.

GN F56A11.4

OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
 CC Rhabditina; Rhabditioidea; Rhabditidae; Pelodierinae; Caenorhabditis.

RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE; 94150718.

RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
 RA BONFIELD J., BORTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
 RA CRAXTON A., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
 RA GARDNER A., GREEN P., HARKINS J., HILLIER J., JIER M., JOHNSTON L.,
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERRY C., RITKEN L., ROOPER A., SANDERS D., SHOWNKEEN R.,
 RA SMALTON N., SMITH A., SONNHAMMER E., STADEN R., STURTON J.,
 RA THIERY-MIGG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.,
 RT 2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans.
 RL Nature 368:32-38(1994).

RL [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA GATUNG S., GOELA D., HARPER M.;
 RL Submitted (DEC-1997) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA WATERSTON R.;
 RL Submitted (DEC-1997) to the EMBL/Genbank/DBJ databases.
 SO SEQUENCE 494 AA; 56780 MW; F59CC020 CRC32;

Query Match 5.5%; Score 87; DB 5; Length 494;
 Best Local Similarity 20.2%; Pred. No. 0.45;
 Matches 51; Conservative 40; Mismatches 96; Indels 66; Gaps 10;

QY 36 ETRATILDHPLRSCNLTLMANVSYVGLPCSVGFTLIQWVVPMDVITRRELVIIRK 95
 DB 53 EINDTCAEDP---PQADAFRMLIYITIGTVCSIGI-----VINTFLILSLRLDVFRS 103
 QY 96 CMHVCILCCANIDIMTS---MMIHG-----YESMALCHCSPGSLQCIAGGQ 139
 DB 104 NIIYFLIACIDIVELCEMILFASLVMDYFRELITTCHEFYIKYVS-----TVGQ 156
 QY 140 VLASTFRVVDGAFNORFIYREVYNNMKREVMNS---SVFMGRHLLIYRLMYD 194
 DB 157 VLASTSTLIYVAAISERYICSLKSSIQFSPQRFLFISIVGACALFKKGSVFPELEL--- 214
 QY 195 GHVGSVPMSPGYSALHGLINNIYVLCSCADLSLRVCCARTRLMRAVRIIA 254
 DB 214 -----OSLPKCPFQNLRL-----DLSEI-----TRTKYLIYGLKLIIS 248

QY 255 EE-TTAMLYSCRT 266
 DB 249 ESKYTMFMCRS 261

RESULT 8

AC 067148 PRELIMINARY; PRT; 211 AA.
 ID 067148

DT 01-AUG-1998 (Tremblrel. 07, Created)
 DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)
 DT 01-NOV-1998 (Tremblrel. 08, Last annotation update)
 DE FORMATE DEHYDROGENASE GAMMA SUBUNIT.

GN Agulifex aeolicus.
 OS Bacteria; Aquificales; Aquificaceae; Agulifex.
 CC [1]

RP SEQUENCE FROM N.A.
 RC STRAIN-VF5;
 RX MEDLINE; 98196666.

RA DECKERT G., WARREN P.V., GASTERLAND T., YOUNG W.G., LENOX A.L.,
 RA GRAHAM D.E., OVERBEK R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,
 RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;
 RT "The complete genome of the hyperthermophilic bacterium Agulifex
 RT aeolicus."
 RL Nature 392:353-358(1998).

RL [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-VF5;
 RA DECKERT G., WARREN P.V., GASTERLAND T., YOUNG W.G., LENOX A.L.,
 RA GRAHAM D.E., OVERBEK R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,
 RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;
 RL Submitted (JUL-1997) to the EMBL/Genbank/DBJ databases.
 DR EMBL; A8000720; AF007109.1;
 SO SEQUENCE 211 AA; 24297 MW; 39A33D85 CRC32;

Query Match 5.4%; Score 85.5; DB 2; Length 211;
 Best Local Similarity 23.4%; Pred. No. 0.25;
 Matches 36; Conservative 20; Mismatches 35; Indels 63; Gaps 7;

QY 64 GLPCSVGFTLIQWVVPMDVITRRELVIIRK-----M 97
 DB 67 GYVFSIGAVIM---ILKWRDVTREDIVMLTKAYISGRHLEPVGKYNAGOLIXA 123
 QY 98 HVCILCCANIDIMTSMTI-----HGYSMALCHC-----CSSPGS 131
 DB 124 WVFTLSAIVFELTGLIMFEPENFSGLVMSVTHEITFIINGAFTIHIYMATVGPVS 183
 QY 132 LQCIAGGQVLASFRVVDGAFNORFIYREVY 165
 DB 184 IMGMIKGVKVSATW-----AKFHHP-KVYREVY 209

RESULT 9

AC 097180 PRELIMINARY; PRT; 954 AA.
 ID 097180

DT 01-MAY-1999 (Tremblrel. 10, Created)
 DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
 DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
 DE RASGAP PROTEIN.

GN RASGAP.
 OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 CC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 CC Ephydroidea; Drosophilidae; Drosophila.
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN-OREGON-R;
 RX MEDLINE; 99147028.

RA FELDMAN P., EICHER E.N., LEWERS S.J., HAFEN E., HUGHES D.A.;
 RT "Control of growth and differentiation by Drosophila RasGAP, a homolog

RT of p120 Ras-GTPase-activating protein.
 RL Mol. Cell. Biol. 19:1928-1937 (1999).
 DR EMBL: AJ012609; CAI0073.1;
 DR HSSP: P20936; 1MER.
 DR PROSITE: PS00509; RAS-GTPASE_ACTIV_1;
 SQ SEQUENCE 954 AA; 107950 MW; 0DB2B774 CRC32;

Query Match 5.48; Score 85.5; DB 5; Length 954;
 Best Local Similarity 22.3%; Pred. No. 1.4;
 Matches 59; Conservative 37; Mismatches 85; Indels 83; Gaps 14;

63 RGPCSVGFLI---QEWVVP-----DVLTRRLVI---LRKCMVCLCANIDI 108
 454 RALPCLATVYLCAPOSESYEMINSKACQDSQSLSAQKVSRLR-----LRCLN--- 506
 109 MTSMMIHGYESMAL-----HCRCSPGSLQCIAGGOVLASWFRVVDGAMFNORFIWYR 162
 506 -----LHVLLEAHLRPFKLVPHYCS-----ISLQVKYKTRVKI-----APEPWEE 548
 163 EVVYNNPKKEMVMSVFM-----RGRHLIYLRL-----WYDGHVGVVPA 203
 549 EFVLDVPPDVVSLLTILISGKRGKSEVAELTIDLSLKNQETEGW--QITGTFM 606
 204 MSFGYSLHCGILNNIVLCCSYCADLSEIRVRCARTRLMDRAVRIAE----- 256
 607 GEMSLRLRMVYLDLIMPCERY--SPLQQLLES-----ELVAVKALALCHNDRV 657
 256 ETAMVLSCTERRRRROFIRLLQ 279
 658 LATALRVFROKRETELIRMLCQ 681

RESULT 10

023017 PRELIMINARY; PRT; 431 AA.

AC 023017;

DT 01-JAN-1998 (TREMBLrel. 05, Created)

DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)

DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)

DE TIG11.7 PROTEIN.

GN Arabidopsis thaliana (Mouse-ear cress).

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC eudicotyledons; Spermatophyta; Magnoliophyta; eudicotyledons;

OC core eudicots; Rosidae; eustoids II; Brassicales; Brassicaceae;

OC Arabidopsis.

RC SEQUENCE FROM N.A.

RA STRAIN-CV. COLUMBIA;

RA OSBORNE B.I., VYSOTSKAIA V.S., TORIUMI M., YU G., OJI O., LIU S.,

RA LI J., HONG L., BUEHLER E., CONWAY A.B., CONWAY A.R., DEMAR K.,

RA FENG J., KIM C., KURZ D., LI Y., SHINN P., SUN H., DAVIS R.W.,

RA ECKERT J.R., FEDERSPIEL N.A., THEOLOGIS A.,

RA Submitted (Aug-1997) to the EMBL/Genbank/DBJ databases.

RA Submitted (Jul-1997) to the EMBL/Genbank/DBJ databases.

RA THEOLOGIS A.;

RA STRAIN-CV. COLUMBIA;

RA THEOLOGIS A.;

RA Submitted (Oct-1997) to the EMBL/Genbank/DBJ databases.

DR EMBL: AC002376; AAB80622.1;

DR MENDEL, 25798; Arath:1088;25798.

DR PFAM: PF00544; pec_lyase; 1;

DR PRINTS: PRO0807; AMBALIENEN.
 SQ SEQUENCE 431 AA; 47770 MW; 868EC285 CRC32;

Query Match 5.3%; Score 84.5; DB 10; Length 431;
 Best Local Similarity 21.7%; Pred. No. 0.7;
 Matches 43; Conservative 35; Mismatches 73; Indels 47; Gaps 12;

43 EDHPLPECNLTLMHNSVYRGVSGFTLQEWV--PMDVLTRELVLRKC----- 97
 127 DNPVNPFRGTL-RAVADRL-----WVFRKDVQLKQELVNSFKTIDG 174
 97 ----MHVCL-CCANIDMTSMIHGYESMALHCRCSPGSLQCIAGGOVLASWFRVVDG 151
 175 RGVNVIANGGCTIIFVTVIVHG-----LHIHCKPFGNMAVRSSETHFGW-RTMADG 228
 152 ----AMNORFIYREVNN-----MPKRVMSVFMGRHLI-VLRMYDGHVGS 199
 229 DAISITGSSHW-----IDNSLSHCADGLVDVAVMSTATISNNLTHNNEVLLGHSDS 284
 200 VV--PAMSEFGYSLHCGI 215
 285 YMRDKAMQVTVIANNHGV 302

RESULT 11

09W6A8 PRELIMINARY; PRT; 354 AA.

AC 09W6A8;

DT 01-NOV-1999 (TREMBLrel. 12, Created)

DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)

DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)

DE BLUE-SENSITIVE OPSIN.

GN Brachydanio rerio (Zebrafish) (Zebra danio).

OS Eukaryota; Metazoa; Chordata; Vertebrata; Actinopterygii;

OC Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;

OC Cyprinidae; Cyprinidae; Rasbortinae; Danio.

RC SEQUENCE FROM N.A.

RA WHITEHEAD T.S., DOHO C.J., HYDE D.R.;

RA Cloning and characterization of six zebrafish photoreceptor opsin

RT cDNAs and immunolocalization of their corresponding proteins."

RL V.S. Neurosci. 0:0-0(1999).

DR EMBL: AF109372; AAD24753.1;

DR PROSITE: PS00238; OPSIN; 1.

SQ SEQUENCE 354 AA; 39483 MW; 535F4C88 CRC32;

Query Match 5.2%; Score 83; DB 13; Length 354;

Best Local Similarity 24.1%; Pred. No. 0.81; 70; Indels 38; Gaps 6;

Matches 40; Conservative 18; Mismatches 85; Indels 38; Gaps 6;

67 CSV-GFT-LIQEWVPMVMDVLTRELVLRKCCHVCLCANIDIMTSMIHGYESMALHC 124
 117 CKIGFISTIGGMSWSLA-----VVALEKWLIVICAPLGFETKTPHALAG----- 164
 125 HCSPGSLQCIAGGOVLASWFRVVDGAMFNORFIWYREVNNPKREVFM----- 177
 164 -CLIPMCWALAGLPPLIGMSRYIPEGLOCSGPDWTTNNKFNNESYVMEFLCFCAVP 222
 177 --SSVEMGRHLIYLRL-----WYDGHVGVVPAVSFGY 208
 223 FSTIVFCYGLLITLKLAKAADSASIQKREKREYTKMAYVYVYGF 268

RESULT 12

023666 PRELIMINARY; PRT; 181 AA.

AC 023666;

DT 01-JAN-1998 (TREMBLrel. 05, Created)

DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)

DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)

DE PUTATIVE PECTATE LYASE (FRAGMENT).
 GN A1.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
 OC Arabidopsis.
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 97422403.
 RA KULIKAVSKAS R., MCCORMICK S.;
 RT "Identification of the tobacco and Arabidopsis homologues of the
 RL pollen-expressed LAT59 gene of tomato."
 DR EMBL: U83620; AAB69760.1; -
 DR MENDEL: 25560; Arabid1088;25560.
 DR PFM: PFM0544; pec_lyase; 1.
 KW Lyase.
 FT NON_TER
 SQ SEQUENCE 181 AA; 20088 MW; EF3426C6 CRC32;

Query Match 5.2%; Score 82.5; DB 10; Length 181;
 Best Local Similarity 20.8%; Pred. No. 0.43;
 Matches 35; Conservative 34; Mismatches 54; Indels 45; Gaps 11;

OY 43 BDHPLPCNTLTMHNSYVGLPCSVGFTLQEWV-PMDVLTRELYLRKC----- 97
 DB 34 DDNPNPGRGL-RHAYIQDRPL-----WIFKRMVTLQELIVSEKTIIDG 81
 OY 97 ---MHVCL-CCANIDITSMHIGESWALHCHCSPGSLQCIAGGOVLASFRMAYDG 151
 DB 82 RGANVHANGCITIOFTNTYVHG-----LHIDCKPGNAM-VRSETHFGW-RIMADG 134
 OY 152 ---AMFNORFIWREVVN-----MPKEVMSVFRGHLTY 188
 DB 135 DAISFGSSHW-----IDHNSLHCHADGLVERSGSTAITISNNLTH 178
 RESULT 13
 O24416 PRELIMINARY; PRT; 405 AA.
 AC 024416;
 DT 01-JAN-1998 (TRENBLREL. 05, Created)
 DT 01-JAN-1998 (TRENBLREL. 05, Last sequence update)
 DT 01-NOV-1999 (TRENBLREL. 12, Last annotation update)
 DE PEPTATE LYASE.
 GN PL.
 OS Fragaria ananassa (Strawberry).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 OC core eudicots; Rosidae; eurosids I; Rosales; Rosaceae; Fragaria.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN-CHANDLER;
 RX MEDLINE: 97435972.
 RA MEDINA-ESCOBAR N., CARDENAS J., MOTANO E., CABALLERO J.L.,
 RA MONO-BLANCO J.;
 RT "Cloning, molecular characterization and expression pattern of a
 RT strawberry ripening-specific cDNA with sequence homology to pectate
 RT lyase from higher plants."
 RL Plant Mol. Biol. 34:867-877(1997).
 DR EMBL: U63550; AAB71208.1; -
 DR MENDEL: 26059; Arab1088;26059.
 DR PFM: PFM0544; pec_lyase; 1.
 KW Lyase.
 FT NON_TER
 SQ SEQUENCE 405 AA; 45744 MW; 2B9944A0 CRC32;

Query Match 5.2%; Score 82.5; DB 10; Length 405;
 Best Local Similarity 24.4%; Pred. No. 1.1;
 Matches 47; Conservative 25; Mismatches 66; Indels 55; Gaps 12;

OY 44 DHPPLPCNTLTMHNSYVGLPCSVGFTLQEWV-PMDVLTRELYLRKC----- 97
 DB 148 DDEVNPPGRGL-RHAYIQDRPL-----WIFKRMVTLQELIVSEKTIIDR 195
 OY 97 ---MHVCL-CCANIDITSMHIGESWALHCHCSPGSLQCIAGGOVLASFRMAYDG- 152
 DB 196 GVNHTAYGGCITIOFTNTYVHG-----LHIDCKPGNAMVRSSPHYGM-RIMADGD 249
 OY 152 ---AMFNORFIWREVVNMPKEVMSVFRGHLTYLRMYGHVGSVPMPSEYS 209
 DB 250 GISFGSSHW-----VDHN-----SLNSCAGGLI-----DANGSTAITISNNYF 290
 OY 210 ALHCGLLNNIVYL 222
 DB 291 THH-----NEWMILL 299

RESULT 14
 O21884 PRELIMINARY; PRT; 1603 AA.
 AC 021884;
 DT 01-NOV-1996 (TRENBLREL. 01, Created)
 DT 01-NOV-1996 (TRENBLREL. 01, Last sequence update)
 DT 01-NOV-1999 (TRENBLREL. 12, Last annotation update)
 DE R09H10.5 PROTEIN.
 GN R09H10.5.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Secernentea; Rhabditiida; Rhabditiida;
 OC Rhabditiida; Rhabditiida; Rhabditiida; Peloderiinae; Caenorhabditis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA BASHAM V.;
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 94150718.
 RA WILSON R., AINSFORTH R., ANDERSON K., BAYNES C., BERKS M.,
 RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
 RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
 RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JER M., JOHNSTON L.,
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIFKIN L., ROOPER A., SAUNDERS D., SHONKKEEN R.,
 RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., STURTON J.,
 RA THIERREY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLIDMAN P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans."
 RL Nature 368:32-38(1994).
 DR EMBL: Z77134; CAB00874.1; -
 DR HSSP: P00750; 1TPG.
 SQ SEQUENCE 1603 AA; 175349 MW; 38FEA309 CRC32;

Query Match 5.1%; Score 82; DB 5; Length 1603;
 Best Local Similarity 21.9%; Pred. No. 5.7;
 Matches 44; Conservative 29; Mismatches 82; Indels 46; Gaps 11;
 OY 9 GMTLRPRSRISRRTPYSNDRLPFETETATLIEDHPLPCNTLTMHNSYV-----R 63
 DB 972 GFSPDPSPDFSRDLVYATHQGVHIL-----SSLQSDPSIDVTSYDNNNNIAFTSLFTR 1028
 OY 64 GLPCSVGFTLQEWV-----VPMVMTRELYLRKCCHVYCC-CCANIDITM---T 110
 DB 1029 TSGCKEYIINNOYTCQASGDTFTMETTTITTEVTLMTORAYCSDLIGCINGIMLGN 1088
 OY 111 SMHIGESWALHCH---CSPGSL-----OC---TAGGOVLASFRMAYDGAMENORITW 160
 DB 1089 CQCVNGYTS--LHCEVPTQONGSVYDFKQCPSTYDG-----DSCQYTTCTWT 1135
 OY 161 YREVVNMPKEVMSVFM 181

DB 1136 --NFETHDPREYNFOQIVFV 1154

RESULT 15
097285 PRELIMINARY; PRT; 433 AA.
AC 097285;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE PFC0915W PROTEIN.
GN PFC0915W.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-3D7;
RA OLIVER K., BOWMAN S., HARRIS D., LAWSON D., QUAIL M., BARRELL B.;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL034559; CAB39031.1;
DR PROSITE; PS00039; DEAD_ATP_HELICASE; 1.
KW ATP-binding; RNA-binding; Helicase.
SQ SEQUENCE 433 AA; 49413 MW; C419B9EF CRC32;

Query Match 5.1%; Score 80.5; DB 5; Length 433;
Best Local Similarity 21.1%; Pred. No. 1.9;
Matches 58; Conservative 49; Mismatches 89; Indels 79; Gaps 15;

QY 46 PLPECTLMHNSVYRGIPCSVGFLLIGEMVPMVLTRELVTLRGMHYCLCAN 105
DB 117 PLLEKCT---HNFTQGL-----ILPTRELALQTSAMIKELGKHKVOCMV 161
QY 106 IDMTSM--MIGYESMALHCHCSPSLQINGGV--LASWFRVVDGA--MENQRF 158
DB 162 ITGISTLRDIDMLN--VHILGTPRIIDLANKDVANLSGCHIMVMDADKLSPEF 219
QY 159 IWTREVVNMPKE--VMMSVVF--MRGNHLYTLRLMTDGHVGSVPAMSF----GYS 209
DB 220 QPIVEELMKFLPKKQIIMTSATFPVTKERRAYLS--DAHEINIMDELTLNGITQY 276
QY 210 A-----LHC-----GIINNTIVLCCSYCADLSEIRVRCARTRRLMLRAVRITAE 255
DB 277 AFWKROKVAHCLNTLFAKLOINQAIIFCNS-----ITREVELAK 315
QY 256 ETTAMLYSCRTERRRQOFIRA-LLOHHRPILMHDY 289
DB 316 KITELGYS-----SFYIHARMSQTHRNRYFHD 343

Search completed: June 23, 2000, 10:07:44
Job time: 947 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2000, 10:05:43 ; Search time 61.02 Seconds
(without alignments)
109.534 Million cell updates/sec

Title: US-09-214-478-4

Sequence: 1 MVLPAIPAPVCDSONECVG.....SIMHDTTNTISVSAHSVQ 114

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 168808 seqs, 58629743 residues

Total number of hits satisfying chosen parameters: 168808

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

1: p1r1.*
2: p1r2.*
3: p1r3.*
4: p1r4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	618	99.5	114	1 Q4ADD2	early E4 13K prote
2	248	39.9	120	2 S10865	early E4 13K prote
3	81	13.0	373	2 F75449	dGTP triphosphohyd
4	80	12.9	118	2 S29651	textilotoxin chain
5	72.5	11.7	1232	2 D64413	cobalamin biosynth
6	67.5	10.9	135	2 S31511	Ig heavy chain - h
7	66	10.6	559	2 S24675	P65 protein - Ectr
8	65.5	10.5	436	2 JC5021	platelet-activatin
9	65	10.5	527	2 S64702	cell surface antiq
10	64.5	10.4	284	2 B70834	hypothetical prote
11	64	10.3	891	2 G75014	valyl-tRNA synthet
12	64	10.3	1083	2 T00790	ubiquitin-specific
13	64	10.3	1695	2 A56921	kinesin family pro
14	63.5	10.2	264	2 T16188	hypothetical prote
15	63.5	10.2	516	2 T40584	probable involveme
16	63.5	10.2	1624	2 C70867	probable Helix-tur
17	63.5	10.2	1639	2 T14181	peptide synthetase
18	63	10.1	247	2 A55717	myelin oligodendro
19	63	10.1	289	2 T35113	diaminopimelate ep
20	63	10.1	488	2 JQ2257	nuclear antigen 21
21	63	10.1	896	2 H70233	hypothetical prote
22	63	10.1	4092	1 S38128	dynein heavy chain
23	62	10.0	211	2 S54092	probable membrane
24	62	10.0	238	2 B47712	myelin oligodendro
25	62	10.0	896	2 D71457	valine--tRNA ligase
26	62	10.0	969	2 T23256	hypothetical prote
27	62	10.0	1759	2 T18868	myoblast city prot
28	61.5	9.9	244	2 T19636	hypothetical prote
29	61.5	9.9	941	2 S09851	hypothetical prote
30	61.5	9.9	1661	2 T31330	head-activator bin

31	61	9.8	213	1 QMYRWT	capsomere protein
32	61	9.8	630	2 S33711	DNA topoisomerase
33	61	9.8	668	1 Q0BEW1	UL52 protein - hum
34	61	9.8	919	1 RNVCZA	DNA-directed RNA p
35	60.5	9.7	210	2 C70627	hypothetical prote
36	60.5	9.7	533	2 S52046	deoxyribodipyrimid
37	60	9.7	114	2 S09829	hypothetical prote
38	60	9.7	121	2 PH1661	Ig heavy chain v r
39	60	9.7	135	2 P10100	Ig heavy chain pre
40	60	9.7	151	2 G70903	hypothetical prote
41	60	9.7	215	1 ASLJSW	vif protein - huma
42	60	9.7	221	2 T15845	hypothetical prote
43	60	9.7	234	2 S60183	protein ExsB - Rhl
44	60	9.7	319	2 T02691	glossyl protein gl
45	60	9.7	397	2 S27600	N-acetylglutamate

ALIGNMENTS

RESULT 1
Q4ADD2
early E4 13K protein - human adenovirus 2
C:Species: Mastadenovirus h2 (human adenovirus 2)
A:Note: host Homo sapiens (man)
C>Date: 02-Apr-1982 #sequence_revision 02-Apr-1982 #text_change 04-Mar-1994
C:Accession: A03806
R:Herisse, J., Rigole, M., Dupont de Dinechin, S., Gallbert, F.
Nucleic Acids Res. 9, 4023-4042, 1981
A:Title: Nucleotide sequence of adenovirus 2 DNA fragment encoding for the carboxylic
A:Reference number: A93733; M01D:82059444
A:Accession: A03806
A:Molecule type: DNA
A:Residues: 1-114 <HER>
A:Note: this probable protein was assigned by correlating EM data and SI digestion st
C:Genetics:
A:Map position: 95.0-96.0
C:Superfamily: adenovirus early E4 13K protein
C:Keywords: early protein

Query Match 99.5% Score 618; DB 1; Length 114;
Best Local Similarity 99.1% Pred. No. 7.1e-60;
Matches 113; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVLPAIPAPVCDSONECVGWLVAYSAVDVIRAAHEGYIEPEARGRLALREWIY 60
Db 1 MVLPAIPAPVCDSONECVGWLVAYSAVDVIRAAHEGYIEPEARGRLALREWIY 60
QY 61 NYTERSKRRDRRRRSYCHARTWCFRCRYDVRSIMHDTTNTISVSAHSVQ 114
Db 61 NYTERAKRRDRRRRSYCHARTWCFRCRYDVRSIMHDTTNTISVSAHSVQ 114
RESULT 2
S10865
early E4 13K protein - human adenovirus 12
C:Species: Mastadenovirus h12 (human adenovirus 12)
C>Date: 21-Nov-1993 #sequence_revision 26-May-1995 #text_change 26-Aug-1999
C:Accession: S10865; S33953
R:Hogenkamp, T.; Esche, H.
Nucleic Acids Res. 18, 3065-3066, 1990
A:Title: Nucleotide sequence of the right 10% of adenovirus type 12 DNA encoding the
A:Reference number: S10860; M01D:90272430
A:Accession: S10865
A:Molecule type: DNA
A:Residues: 1-120 <HOG>
A:Cross-references: EMBL:X51800
A:Note: the authors translated the codon TCG for residue 83 as Tyr
R:Spengler, J.
submitted to the EMBL Data Library, June 1993
A:Reference number: S33928
A:Accession: S33953

A;Molecule type: DNA
A;Residues: 1-120 <SPB>
A;Cross-references: EMBL:X73487; NID:G313361; PIDD:CAAS1902.1; P;ID:G313387
C;Superfamily: adenovirus early E4 13k protein

Query Match	39.98;	Score 248;	DB 2;	Length 120;
Best Local Similarity	40.28;	Pred. No. 8.4e-20;		
Matches 43; Conservative	23;	Mismatches 41;	Indels 0;	Gaps 0

Q7 1 MFLPALPAPPCDSQNECVGMLGYAIVAVDVIRAAAHGVYIEPEARGLDALREWITY 60
| | | | | : | | | | | : | | | | | : | | | | |
Db 1 MPFLCIPPPVSRDTAAICIAMLGIAHASCVDTLRFIKHDKLTKEPAEYILASLREWLYE 60

61 NYTTERSKRDRRRSVCHARTWFCFKYDYVRSIMWHDTTNTISV 107

E75449

[illegible]

A:/title: genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1.
A:/Reference number: A75250
A:/Accession: F75449
A:/Status: preliminary
A:/Molecule type: DNA
A:/Residues: 1-373 <WHI>
A:/Cross-references: GB:AE001952; GB:AE000513; NID:96458725; PIDN:AAF10581.1; PID:9645873
A:/Experimental source: strain R1
C:/Genetics:
A:/gene: DRI006
A:/map position: 1

Query Match	13.0%;	Score 81;	DB 2;	Length 373;
Best Local Similarity	35.3%;	Pred. No. 0.31;		
Matches 24; Conservative	7;	Mismatches 21;	Indels 16;	Gaps 3

QY 59 YNYYTER 66
| :: ||
Db 305 YRHMÖVER 312

Query Match	12.9%;	Score 80;	DB 2;	Length 118;
Best Local Similarity	29.0%;	Pred. No. 0.13;		
Matches 29; Conservative	12;	Mismatches 39;	Indels 20;	Gaps

```

09 12 CDSQNECVGWIGVAY-----SAVVDVIRAA-AHEGVYIEPEARRLDLAREWITY 60
      |::|||::|||::|::|::|
Db 11 CANKKRPGMHIYANTGCYCGSGGRGTPVVDVDRCCQAHDKCIEDAE---KLGCIYKWTYY 67

```

```

QY 61 NYTTERS---RRDRRRRSVCHARTWF--CFPKYDIYRR 94
      ||| : | : : ||| : | | | |
DQ 68 NYGGANGPYCATRTKQRFVNCDCVVAADCFASYPYRRR 107

```

RESULT 5
D64413
cobalamin biosynthesis protein N homolog - Methanococcus jannaschii
C:Species: Methanococcus jannaschii

Query Match	11.7%;	Score 72.5;	DB 2;	Length 1232;
Best Local Similarity	23.8%;	Pred. No. 8.5;		
Matches 35; Conservative	14;	Mismatches 51;	Indels 47;	Gaps 6

QY 3 LPALPAPVC-----DSQNECVGMLGVASAV----- 30
 :| | | : | : | : | | |
 Db 86 IPPLPPLTISEIHDDLEKDRIVDDYKNNKYVYKLG--IEGVYNNKLLLYIPANRFGNTLV 143

```

      : |          : ||| : |   || | | | : | : | | | |
-----+-----+-----+-----+-----+-----+-----+
Db    144 EYEEPRMPWQGIYYKGKYETLDD---YINYLKELGRDLDRPIIGVLEFYRNWFANNN 198
```

Dy 89 YDVRKRSIMHDTTIN-----TISVSASAH 111
||| | | |||
Db 199 IDYVNDLT--DIENKGALPIAVFSSH 223

RESULT 6
S31511
ig heavy chain - human
C.Species: Homo sapiens (man)
C.Date: 13-Jan-1995 #sequence
C.Accession: S31511
R.Chastagner, P.; Demaison, C.; Theze, J.; Zouali, M.
submitted to the EMBL Data Library, December 1992
A.Description: Dominance of clonotypic patterns and variable gene usage of anti-DNA-a
A.Reference number: S31509
A.Accession: S31511
A.Status: preliminary
A.Molecule type: mRNA
A.Residues: 1-155 <CH>
A.Cross-references: EMBL:X69866; NID:g33094; PIDD:CAA49500.1; PID:g33095
C.Superfamily: immunoglobulin V region, immunoglobulin homology
C.Keywords: heterotetramer; immunoglobulin
P:47-129/Domain: immunoglobulin homology <IM>

Query Match 10.4%; Score 64.5; DB 2; Length 284;
Best Local Similarity 30.4%; Pred. No. 15;
Matches 17; Conservative 9; Mismatches 21; Indels 9; Gaps 2;

OY 63 YTEBSKRRDRRRSCHARTWCFKRYDYRRST-----WHDTTNTISVSASVS 113
DB 21 YVDIDRRRRRRR-----RKSMAHSGHGFYERSTELIKRWGVSATGDAVAHNV 72

RESULT 11

G75014

valyl-tRNA synthetase (vals) PAB125 - Pyrococcus abyssi (strain Orsay)

C:Species: Pyrococcus abyssi

C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 24-Nov-1999

C:Accession: G75014

R:Anonymous; Genoscope

submitted to the EMBL Data Library, July 1999

A:Description: Pyrococcus abyssi genome sequence: Insights into archaeal chromosome stru

A:Reference number: A75001

A:Accession: G75014

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-891 <KAM>

A:Cross-references: GB:AJ248286; GB:AL096836; NID:95458960; PIDN:CAB50557.1; PID:el51645

A:Experimental source: Strain Orsay

C:Genetics:

A:Gene: PAB1255

C:Superfamily: valine-tRNA ligase

Query Match 10.3%; Score 64; DB 2; Length 891;
Best Local Similarity 23.1%; Pred. No. 51;
Matches 25; Conservative 16; Mismatches 43; Indels 24; Gaps 5;

OY 4 PALPAPVCDSONECVGV-----WLGVAVSAYVDVIRAAHGGVYIEPEARGRLDALREMI 58
DB 435 PVEKCPVCGAIEIPYDVLDCWD-----SITPLILTRMHAIAGDEA-----KKWF 484

OY 59 YNNTYERSKRRDRRRRSVCHARTW---FCFRKYDYVRSIWMHTTN 103
DB 485 EHNPTA-----LRFOTDIIITWAFYIIRFKLGGKRWKNOIVIN 526

RESULT 12

T00790

ubiquitin-specific proteinase homolog F24L7.8 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cross)

C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 30-Apr-1999

C:Accession: T00790

R:Rounsailey, S.D.; Lih, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul

submitted to the EMBL Data Library, February 1998

A:Description: Arabidopsis thaliana chromosome II BAC F24L7 genomic sequence.

A:Reference number: Z14204

A:Accession: T00790

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1083 <KROD>

A:Cross-references: EMBL:AC003974; NID:92914688; PID:92914695

A:Experimental source: cultivar Columbia

C:Genetics:

A:Map-position: 2

A:Introns: 416/2; 1004/2

A:Note: F24L7.8

Query Match 10.3%; Score 64; DB 2; Length 1083;
Best Local Similarity 24.3%; Pred. No. 62;
Matches 18; Conservative 12; Mismatches 30; Indels 14; Gaps 1;

OY 14 SQNECVGLVAVSAVVDVIRAAHGGVYI-----EPEARGRLDALREMI 59
DB 931 SDNECVGLVAVSAVVDVIRAAHGGVYI-----EPEARGRLDALREMI 59

OY 931 SDNECVGLVAVSAVVDVIRAAHGGVYI-----EPEARGRLDALREMI 59
DB 931 SDNECVGLVAVSAVVDVIRAAHGGVYI-----EPEARGRLDALREMI 59

OY 60 YNNTYERSKRRDR 73
DB 991 FQEPIDLSKYMDTR 1004

RESULT 13

A56921

kinesin family protein KIF1A - mouse

C:Species: Mus musculus (house mouse)

C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 13-Aug-1999

C:Accession: A56921

R:Okada, Y.; Yamazaki, H.; Sekine-Aizawa, Y.; Hirokawa, N.

Cell 81, 769-780, 1995

A:Title: The neuron-specific kinesin superfamily protein KIF1A is a unique monomeric

A:Reference number: A56921; M01D:95292344

A:Accession: A56921

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1695 <RES>

A:Cross-references: GB:D29951; NID:9976234; PIDN:BA06221.1; PID:9976235

A:Superfamily: kinesin-related protein unc-104; kinesin motor domain homology; plects

C:Keywords: P-loop

F/6-360/Domain: kinesin motor domain homology <KMO7>

F/97-104/Region: nucleotide-binding motif A (P-loop)

Query Match 10.3%; Score 64; DB 2; Length 1695;
Best Local Similarity 23.4%; Pred. No. 97;
Matches 32; Conservative 19; Mismatches 56; Indels 30; Gaps 7;

OY 7 PAPPV--CDSONECVGLV-----AYSADVIRAAHGGVYIEPEAR---GRLD 52
DB 613 PAPVDMARQELLEKGIDMKOMEGRLOLEQYREREATYLLQORLDYESTLE 672

OY 53 ALREYIYTYERSKRRDRRRSV-----CHARTWCFKRYDY-----VRSIMHTD 101
DB 673 ALQKMDRSYYPEVNEEPEDEVOQTERECALM-AFRKWKYQVTSRLDLNGVAI 731

OY 101 ---TNTISVSASVS 114
DB 732 FLKEMALSVELKKRVO 748

RESULT 14

T16188

hypothetical protein F27D9.6 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999

C:Accession: T16188

R:Bentley, D.

submitted to the EMBL Data Library, February 1996

A:Description: The sequence of C. elegans cosmid F27D9.

A:Reference number: Z18473

A:Accession: T16188

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-264 <BEN>

A:Cross-references: EMBL:U49829; NID:91203924; PID:91203928; PIDN:AAA93385.1; CESP:F2

C:Genetics:

A:Gene: CESP:F27D9.6

A:Introns: 28/2; 57/3; 117/2; 173/3; 231/2

Query Match 10.2%; Score 63.5; DB 2; Length 264;
Best Local Similarity 23.2%; Pred. No. 17;
Matches 26; Conservative 17; Mismatches 48; Indels 21; Gaps 6;

OY 13 DSQNECV---GW---LGVAVSAYVDVIRAAHGGVYIEPEARGRLDALREMIYNYT 64
DB 6 DGHYCVGSINGWGEWGLSCTSKFAVNGAM-ESLQMLRDRG-LEGIKITTLTYFYFA 63

OY 65 ESKRRDRRRRSVCHARTWCF-----KKYDYVRSIWMHTTNTISV 108
DB 65 ESKRRDRRRRSVCHARTWCF-----KKYDYVRSIWMHTTNTISV 108

Db 64 RTPMLENNMRPTC--TWPEPMISIRSCSKMVDISILKEKHAIVPSYITLI 112

RESULT 15

T40584 Probable involvement in ergosterol biosynthesis - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999

C:Accession: T40584

R:Seeger, K.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, January 1999

A:Reference number: 221938

A:Accession: T40584
A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-516 <SEE>
A:Cross-references: EMBL:AL035216; PIDN:CNA22812.1; GSPDB:GN00067; SPDB:SPBC646.08c

A:Experimental source: strain 972h-; cosmid c646

C:Genetics: SPDB:SPBC646.08c

A:Map position: 2

Query Match

Best Local Similarity 22.9%; Pred. No. 34;
Matches 25; Conservative 14; Mismatches 47; Indels 23; Gaps 4;

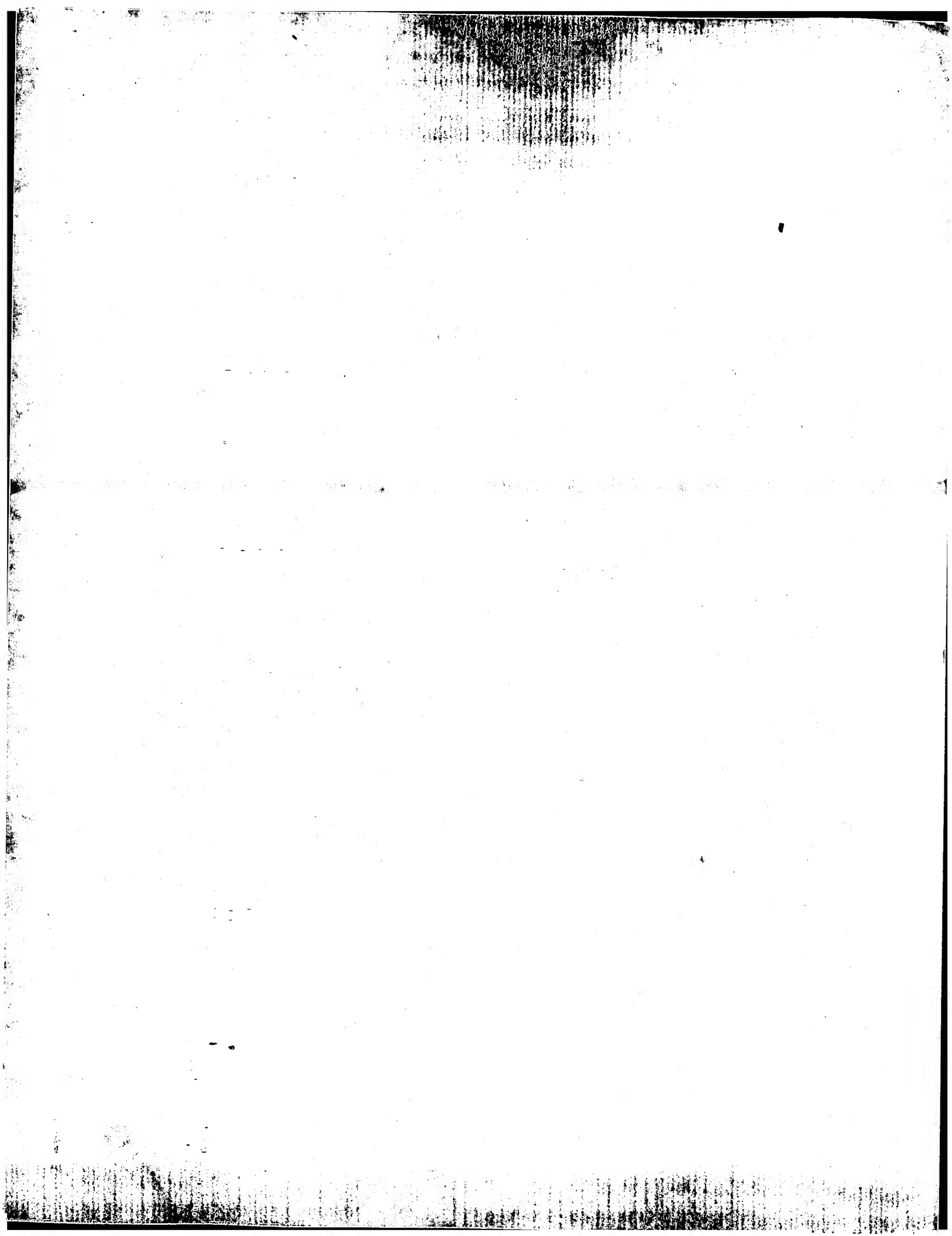
21 WLGVASAV-----VDVRAAAHGGVYTEPEARGRLDALREMYIYYNTERS 67

Db 354 WLKPRSLVEGVGYRYDPNDIDISIKAVKENILATFKGN-----WRNCIFYSAGESE 408

68 KRR--DRRRSVCHARTWCFGRKYDVRRSIVHDTTNTISVSAHSVQ 114

Db 409 SRRLVDLNLDELVKRCPPIDKOPFESRKIMPPVTHN---ILAKHYIQ 454

Search completed: June 23, 2000, 10:05:47
Job time: 1752 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2000, 10:14:48 ; Search time 13.99 Seconds
(without alignments)

248.168 Million cell updates/sec

Title: US-09-214-478-4

Perfect score: 621
Sequence: 1 MYLPALPAPVCDSONECVGV.....SIWHDTTNTISVSAHSVQ 114

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 83857 seqs, 30454973 residues

Total number of hits satisfying chosen parameters: 83857

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : SwissProt_38.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	618	99.5	E413_ADE02	P03240 human adeno
2	248	39.9	E413_ADE12	P36709 human adeno
3	80	12.9	PA2A_PSETE	P23026 pseudonata
4	68	11.0	RP54_RHIME	P17263 rhizobium m
5	67	10.8	Y011_BPL2	P42346 bacteriopho
6	66	10.6	RRPL_AHSV9	O70695 african hor
7	65.5	10.5	PAPA_CAVPO	P70683 C platelet-
8	64	10.3	KFLA_HUMAN	Q12756 homo sapien
9	64	10.3	KFLA_MOUSE	P33173 mus musculu
10	63	10.1	MOG_MOUSE	O61885 mus musculu
11	63	10.1	DAPF_STRCO	O69869 streptomyc
12	63	10.1	CEGT_HUMAN	Q16739 homo sapien
13	63	10.1	PSD3_DAVCA	O06364 danus caro
14	63	10.1	DYHC_YEAST	P36022 saccharomyc
15	62	10.0	MOG_RAT	O63345 rattus norv
16	62	10.0	EXG_YAPLI	Q12725 yarrowia 11
17	62	10.0	STAR_DROME	P42519 drosophila
18	61.5	9.9	UL87_HCMVA	P16730 human cytom
19	61	9.8	VP9_MTV	P31326 wound tumor
20	61	9.8	PARF_SALTY	P31398 salmonella
21	61	9.8	UL52_HCMVA	P16793 human cytom
22	61	9.8	RPO2_CAVPK	P16716 capripoxvir
23	60.5	9.7	Y4YR_RHISN	P55726 rhizobium s
24	60	9.7	UL66_HCMVA	P16822 human cytom
25	60	9.7	VIF_HV2ST	P20878 human immun
26	60	9.7	ARGA_PSEAE	P22567 pseudomonas
27	60	9.7	COBL_PSEAE	P21921 pseudomonas
28	59.5	9.6	LCB3_ROBPS	O41160 scobia pse
29	59.5	9.6	POLL_SCTCO	O03277 sclaria copr
30	59.5	9.6	SYL_TREPA	O83466 treponema p
31	59.5	9.6	YLJ2_CAEEL	P34367 caenorhabdl
32	59	9.5	HLIC_HUMAN	P01744 homo sapien
33	59	9.5	MOG_BOVIN	P55803 bos taurus
34	59	9.5	VP9_MTVNJ	P31611 wound tumor

35	59	9.5	339	1	YGBQ_HAEIN	P44039 haemophilus
36	58.5	9.4	445	1	SPRI_YEAST	P32603 saccharomyc
37	58.5	9.4	833	1	CM41_YEAST	P53008 saccharomyc
38	58	9.3	185	1	VCO7_ADE40	O89532 human adeno
39	58	9.3	215	1	VIF_HV2G1	P18043 human immun
40	58	9.3	271	1	YK23_YEAST	P36136 saccharomyc
41	58	9.3	358	1	YK15_CAEEL	P34316 caenorhabdl
42	58	9.3	426	1	YJIN_ECOLI	P39385 escherichia
43	58	9.3	508	1	COB1_MYCTU	Q10677 mycobacteri
44	57.5	9.3	182	1	LEC7_CAEEL	O09605 caenorhabdl
45	57.5	9.3	251	1	HXB4_FUGRU	O13074 fugu rubrip

ALIGNMENTS

RESULT 1	E413_ADE02	STANDARD:	PRT:	114 AA.
ID	E413_ADE02	STANDARD:	PRT:	114 AA.
AC	P03240;			
DT	21-JUL-1986 (Rel. 01, Created)			
DT	21-JUL-1986 (Rel. 01, Last sequence update)			
DT	01-APR-1990 (Rel. 14, Last annotation update)			
DE	PROBABLE EARLY E4 13 KD PROTEIN.			
OS	Human adenovirus type 2.			
OC	Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE; 82059444.			
RA	Hershe J., Rigole M., Dupont de Dinechin S., Galibert F.,			
RT	"Nucleotide sequence of adenovirus 2 DNA fragment encoding for the			
RI	carboxylic region of the fiber protein and the entire E4 region."			
RL	Nucleic Acids Res. 9:4023-4042(1981).			
CC	-1- MISCELLANEOUS: THIS PROBABLE PROTEIN WAS ASSIGNED BY CORRELATING			
CC	EM DATA AND S1 DIGESTION STUDIES.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; J01917; -, NOT_ANNOTATED_CDS.			
DR	PIR; A03806; Q4AD2.			
KW	Early protein.			
SO	SEQUENCE 114 AA; 13321 MW; 72FB72749D563457 CRC64;			
Query Match 99.5%; Score 618; DB 1; Length 114;				
Best Local Similarity 99.1%; Pred. No. 1e-61;				
Matches 113; Conservative 1; Mismatches 0; Indels 0; Gaps 0;				
OY	1 MYLPALPAPVCDSONECVGVSAVDYIRAAHEGYIEFEANGRLDALREWTYY 60			
DB	1 MYLPALPAPVCDSONECVGVSAVDYIRAAHEGYIEFEANGRLDALREWTYY 60			
OY	61 NYTTERSKRRDRRRNSVCHARTWFCGRKDYVRSIMHDTTNTISVSAHSVQ 114			
DB	61 NYTTERSKRRDRRRNSVCHARTWFCGRKDYVRSIMHDTTNTISVSAHSVQ 114			
RESULT 2				
ID	E413_ADE12	STANDARD:	PRT:	120 AA.
AC	P36709;			
DT	01-JUN-1994 (Rel. 29, Created)			
DT	01-JUN-1994 (Rel. 29, Last sequence update)			
DT	01-JUN-1994 (Rel. 29, Last annotation update)			
DE	PROBABLE EARLY E4 13 KD PROTEIN.			
OS	Human adenovirus type 12.			
OC	Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.			

RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 94076430.
 RA Sprengel J., Schmitz B., Heuss-Neltzel D., Zock C., Doerfler W.;
 RT "Nucleotide sequence of human adenovirus type 12 DNA: comparison
 RL functional analysis.";
 RU J. Virol. 68:379-389(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 90272430.
 RA Hogenkamp T., Esche H.;
 RT "Nucleotide sequence of the right 10% of adenovirus type 12 DNA
 RL encoding the entire region E4.";
 RU Nucleic Acids Res. 18:3065-3066(1990).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X73487; CAAS1902.1;
 DR EMBL: X51800; CAB57853.1;
 DR PIR: S33933; S33933.
 DR PIR: S10865; S10865.
 KW Early Protein.
 SQ SEQUENCE 120 AA; 13600 MW; 2004E4A61B792AF CMC64;
 Query Match 39.9%; Score 248; DB 1; Length 120;
 Best Local Similarity 40.2%; Pred. No. 1e-20;
 Matches 43; Conservative 23; Mismatches 41; Indels 0; Gaps 0;
 Oy 1 MYPLPAPPCVDSQNECGMVGVAIVYDVRRAAHBQYIEFAGRDALREWTYY 60
 Db 1 MFLPCLPPPPVSRDPAACIAMLGLAHASCVDTLRKHKHDKITPEAYTILASIREWLTF 60
 Oy 61 NYTTERSKRRDRRRRSVCHARTWFCFRKYDYRRSINWDTTNTTSY 107
 Db 61 ALTERORCKOKGKRGALTSGRTWFCFFYEDARKSVYDARQVSL 107
 RESULT 3
 PAZA_PSETE STANDARD; PRT; 118 AA.
 AC P23026;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE PHOSPHOLIPASE A2 HOMOLOG, TEXTILOTOXIN A CHAIN (EC 3.1.1.4).
 OS Pseudonaja textilis (Eastern brown snake).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Lepidosaurota;
 OC Squamata; Scleroglossa; Serpentes; Colubroidae; Elapidae;
 OC Acanthophiline; Pseudonaja.
 RN [1]
 RP SEQUENCE.
 OS TISSUE-VENOM.
 RX MEDLINE: 93160241.
 RA Pearson J.A., Tyler M.I., Retson K.V., Howden M.E.H.;
 RT "Studies on the subunit structure of textilotoxin, a potent
 RT presynaptic neurotoxin from the venom of the Australian common brown
 RT snake (Pseudonaja textilis). 3. The complete amino-acid sequences of
 RT all the subunits.";
 RU Blochm. Biophys. Acta 1161:223-229(1993).
 RN [2]
 RP SEQUENCE.
 OS TISSUE-VENOM.
 RX MEDLINE: 88000715.
 RA Tyler M.I., Barnett D., Nicholson P., Spence I., Howden M.E.H.;
 RT "Studies on the subunit structure of textilotoxin, a potent
 RT neurotoxin from the venom of the Australian common brown snake

RT (Pseudonaja textilis).";
 RL Blochm. Biophys. Acta 915:210-216(1987).
 CC -1- FUNCTION: POTENT PRESYNAPTIC NEUROTOXIN POSSESSING PHOSPHOLIPASE
 CC ACTIVITY. SUBUNIT A IS LETHAL TO MICE AT 4 MG/KG (I.V.). IT IS
 CC ESSENTIAL FOR THE NEUROTOXICITY OF TEXTILOTOXIN. SUBUNIT A
 CC POSSESSES A LOW PHOSPHOLIPASE ACTIVITY.
 CC -1- FUNCTION: PA2 CATALYZES THE CALCIUM-DEPENDENT HYDROLYSIS OF THE
 CC -1- 2-ACYL GROUPS IN 3-SN-PHOSPHOGLYCERIDES.
 CC -1- CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + H(2)O = 1-ACYLGLYCERYL-
 CC PHOSPHOCHOLINE + A FATTY ACID ANION.
 CC -1- SUBUNIT: PENTAMER OF THREE NON-COVALENTLY LINKED SUBUNITS A, B,
 CC AND C, AND TWO IDENTICAL COVALENTLY LINKED D SUBUNITS.
 CC -1- MISCELLANEOUS: ALL SUBUNITS ARE NECESSARY FOR MAXIMUM LETHALITY.
 CC LD(50) IS 0.001 MG/KG BY INTRAPERITONEAL INJECTION.
 CC -1- SIMILARITY: BELONGS TO THE PHOSPHOLIPASE A2 FAMILY.
 DR PIR: S29651; S29651.
 DR HSP: P00609; 2NOT.
 DR PRINTS: PR00389; PHPLIPASEA2.
 DR PROSITE: PS00118; PA2_HIS; 1.
 DR PROSITE: PS00119; PA2_ASP; 1.
 DR PRAM: PF00068; Phoslip; 1.
 KW Hydrolase; Lipid degradation; Calcium; Multigene family; Venom;
 FT ACT_SITE 48 48 BY SIMILARITY.
 FT ACT_SITE 92 92 BY SIMILARITY.
 FT DISULFID 11 71 BY SIMILARITY.
 FT DISULFID 27 117 BY SIMILARITY.
 FT DISULFID 29 45 BY SIMILARITY.
 FT DISULFID 44 98 BY SIMILARITY.
 FT DISULFID 51 91 BY SIMILARITY.
 FT DISULFID 60 84 BY SIMILARITY.
 FT DISULFID 78 89 BY SIMILARITY.
 SQ SEQUENCE 118 AA; 13849 MW; 3B43FBC0A751C09E CMC64;
 Query Match 12.9%; Score 80; DB 1; Length 118;
 Best Local Similarity 29.0%; Pred. No. 0.041;
 Matches 29; Conservative 12; Mismatches 39; Indels 20; Gaps 5;
 Oy 12 CDSQNCVGMVGVAIVYDVRRAAHBQYIEFAGRDALREWTYY 60
 Db 11 CANRYRPPGMHANYGCGSGSGRTPVDVDRCCQADHKCYEDAE--KLGCPKPTTY 67
 Oy 61 NYTTERSKRRDRRRRSVCHARTWFCFRKYDYRRSINWDTTNTTSY 94
 Db 68 NYTCGANGPYCKTRTKQCFVNCVVAADGCASTPYNRR 107
 RESULT 4
 RP54_RHIME STANDARD; PRT; 514 AA.
 ID RP54_RHIME
 AC P17263; OS9751;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 15-FEB-2000 (Rel. 39, Last sequence update)
 DT 15-FEB-2000 (Rel. 39, Last annotation update)
 DE RNA POLYMERASE SIGMA-54 FACTOR (VERSION 1).
 GN RPN OR NTRA.
 OS Rhizobium meliloti.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Sinorhizobium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 87222158.
 RA Ronson C.W., Nixon B.T., Albright L.M., Ausubel F.M.;
 RT "Rhizobium meliloti ntrA (rpoA) gene is required for diverse
 RT metabolic functions.";
 RU J. Bacteriol. 169:2424-2431(1987).
 RN [2]
 RP REVISIONS TO C-TERMINUS.
 RX MEDLINE: 89197760.
 RA Albright L.M., Ronson C.W., Nixon B.T., Ausubel F.M.;
 RT "Identification of a gene linked to Rhizobium meliloti ntrA whose
 RT product is homologous to a family to ATP-binding proteins.";

```

CC modified and this statement is not removed. Usage by and for commercial/
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L13696; AAA87967.1; -.
CC DR Hypothetical protein.
CC KM Hypothetical protein.
CC SQ SEQUENCE 218 AA; 25564 MW; A6DD8C582B8EFCPE CRC64;

Query Match 10.8%; Score 67; DB 1; Length 218;
Best Local Similarity 24.3%; Pred. No. 2;
Matches 17; Conservative 12; Mismatches 31; Indels 10; Gaps 2;

QY 14 SQNCEVGLGAVYANVVDVTRAAHAG-----VYIEPEARGR-----LDLAREWYYNY 63
D7 47 SNDELIGMGCSRGLYDAFVRLEREGLIKRIYSDTKRKRTETIEFNIDLAASWLKLTLY 106
QY 64 TERSKRRDR 73
D8 107 DKAYKNAPK 116

RESULT 6
RRPL_AHSV9 STANDARD; PRT; 1305 AA.
AC 070695;
AT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (Vp1).
GN SL.
OS African horse sickness virus 9 (AHSV-9) (African horse sickness virus
OS (serotype 9)).
OC Viruses; dsRNA viruses; Reoviridae; Orbivirus.
OC [1]
RN SEQUENCE FROM N.A.
RX MEDLINE: 98202706.
RA Vreede F.T., Huismans H.;
RT "Sequence analysis of the RNA polymerase gene of African horse
RL sickness virus."
RL Arch. Virol. 143:413-419(1998).
CC -1- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE = N PYROPHOSPHATE
CC + RNA(N).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial/
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U94887; AAC40586.1; -.
CC DR Transferrase; Nucleosidyltransferase; RNA-directed RNA polymerase.
CC KM SEQUENCE 1305 AA; 150293 MW; 321E9EF3C6F11A CRC64;
CC SQ

Query Match 10.6%; Score 66; DB 1; Length 1805;
Best Local Similarity 30.2%; Pred. No. 20;
Matches 19; Conservative 9; Mismatches 21; Indels 14; Gaps 3;

QY 49 GRDLALREIYYNYTERKKRRDRRRRSYCHART--WFCFRKYD-----YVRSI 96
D8 111 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D8 22 GRDGRK--MYEYYRYSKMKRETRRKKTGYKTTDDEFLERDAGRLKLYDLYVIRENS 79
QY 97 WHD 99
D8 80 WED 82

RESULT 7
PAPA_CAVPO STANDARD; PRT; 436 AA.

```

AC 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-FEB-2000 (Rel. 39, Last annotation update)
 DE PLATELET-ACTIVATING FACTOR ACETYLHYDROLASE PRECURSOR (EC 3.1.1.47)
 DE (PAF ACETYLHYDROLASE) (PAF 2-ACTIHYDROLASE) (LID-ASSOCIATED
 DE PHOSPHOLIPASE A2) (LID-PLA(2)) (2-ACETYL-1-ALKYLGLYCEROPHOSPHOCHOLINE
 DE ESTERASE) (1-ALKYL-2-ACETYLGLYCEROPHOSPHOCHOLINE ESTERASE).
 GN PLA2G7.
 OS Cavia porcellus (Guinea pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Rodentia; Hystriocognathi; Caviidae; Cavia.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-HARTLEY; TISSUE-LIVER;
 RC MEDLINE: 97103479.
 RA Karsawa K., Kuge O., Kawasaki K., Nishijima M., Nakano Y., Tomita M.,
 RA Yokoyama K., Setaka M., Nojima S.;
 RT "Cloning, expression and characterization of plasma
 RT platelet-activating factor acetylhydrolase from guinea pig.";
 RT J. Biochem. 120:838-844(1996).
 RL -1- FUNCTION: MODULATES THE ACTION OF PLATELET-ACTIVATING FACTOR (PAF)
 CC BY HYDROLYZING THE SN-2 ESTER BOND TO YIELD THE BIOLOGICALLY
 CC INACTIVE LYSO-PAF. HAS A SPECIFICITY FOR SUBSTRATES WITH A SHORT
 CC RESIDUE AT THE SN-2 POSITION. IT IS INACTIVE AGAINST LONG-
 CC CHAIN PHOSPHOLIPIDS.
 CC -1- CATALYTIC ACTIVITY: 2-ACETYL-1-ALKYL-SN-GLYCERO-3-PHOSPHOCHOLINE +
 CC H(2O) -> 1-ALKYL-SN-GLYCERO-3-PHOSPHOCHOLINE + ACETATE.
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.
 CC -1- TISSUE SPECIFICITY: PLASMA.
 CC -1- SIMILARITY: PARTIAL WITH OTHER LIPASES (PANCREATIC, GASTRIC,
 CC HEPATIC, LINGUAL, LIPOPROTEIN, BACTERIAL, ETC.).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: D67037; BAA11054.1;
 DR PROSITE: PS00120; LIPASE_SER; 1.
 DR KN Hydrolyase: lipid degradation; Glycoprotein; Signal.
 FT SIGNAL 1 21
 FT CHAIN 22 436
 FT ACT_SITE 271 271
 FT ACT_SITE 294 294
 FT ACT_SITE 349 349
 FT CARBOHYD 76 76
 FT CARBOHYD 200 200
 FT CARBOHYD 324 324
 FT SEQUENCE 436 AA; 49062 MW; C359D96E392FEF1 CRC64;

Query Match 10.5%; Score 65.5; DB 1; Length 436;
 Best Local Similarity 28.2%; Pred. No. 6.8;
 Matches 20; Conservative 10; Mismatches 32; Indels 9; Gaps 3;

QY 22 LGVAVSAVDVIRAAHE-----GYIIEPEARGRDLARENTY---NYIERSKRRDR 73
 DB 163 IGIELASHGFIYAANVHEDESAATYTFQDAPALMSGRSMYIYVGLTEFERKRLQ 222
 QY 74 RRSVC-HARTW 83
 DB 223 RGECCQALSW 233

RESULT 8
 ID KFLA_HUMAN STANDARD; PRT; 1690 AA.
 AC 012756;

DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE KINISIN-LIKE PROTEIN KIFIA (AXONAL TRANSPORTER OF SYNAPTIC VESICLES).
 GN KIFIA OR ATSV.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN;
 RC MEDLINE: 96299637.
 RA Furlong R.A., Zhou C.Y., Ferguson-Smith M.A., Affara N.A.;
 RT "Characterization of a kinesin-related gene ATSV, within the tuberous
 RT sclerosis locus (TSC1) candidate region on chromosome 9q34.";
 RT Genomics 33:421-429(1996).
 RL -1- FUNCTION: MOTOR FOR ANTEROGRADE AXONAL TRANSPORT OF SYNAPTIC
 CC VESICLE PRECURSORS (BY SIMILARITY).
 CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. UNC-104
 CC SUBFAMILY.
 CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: X90840; CA62346.1;
 DR HSSP: P17119; 3KAR.
 DR MIM: 601255.
 DR PRINTS: PR00380; KINESINHEAV.
 DR PROSITE: PS00411; KINESIN_MOTOR_DOMAIN1; 1.
 DR PROSITE: PS50067; KINESIN_MOTOR_DOMAIN2; 1.
 DR PROSITE: PS50003; PH_DOMAIN; 1.
 DR PROSITE: PS50006; PH_DOMAIN; 1.
 DR PFAM: PF00169; PH; 1.
 DR PFAM: PF00225; kinesin; 1.
 DR PFAM: PF00498; FHA; 1.
 DR Motor protein; Microtubules; ATP-binding; Coiled coil.
 FT DOMAIN 1 361
 FT DOMAIN 366 383
 FT DOMAIN 429 462
 FT DOMAIN 516 572
 FT DOMAIN 622 681
 FT DOMAIN 801 822
 FT DOMAIN 1575 1673
 FT NP_BIND 97 104
 FT SEQUENCE 1690 AA; 191083 MW; D8DDE0784624FBAD CRC64;

Query Match 10.3%; Score 64; DB 1; Length 1690;
 Best Local Similarity 23.4%; Pred. No. 43;
 Matches 32; Conservative 19; Mismatches 56; Indels 30; Gaps 7;

QY 7 PAPVY-CDSONECVGLGV-----ASAVVDVIRAAHGVYIEPAR-----GRD 52
 DB 613 PAEVDYAFQRELLEQOGIDMKQENORLQLEDYRREREATYLLQQRIDYRSKLE 672
 QY 53 ALREMIYNYIYTERSKRRDRRRRSV-----CHARTWCFRKDY-----VRSIWHDT 101
 DB 673 ALQKQMSRYIYVNEEEPEDEYQWTERECCLALM-AFRKKWKQFISRLDINGNAI 731
 QY 101 ----TNTIYVSAHSVQ 114
 DB 732 FLKEANAISVELKKVQ 748

RESULT 9

ID	NAME	STANDARD	PRT	1695 AA
AC	PF33173	061770		
DT	01-OCT-1993	(Rel. 27, Created)		
DT	01-NOV-1997	(Rel. 35, Last sequence update)		
DT	15-JUL-1998	(Rel. 36, Last annotation update)		
DE	KINESIN-LIKE PROTEIN KIF1A.			
GN	KIF1A OR KIF1.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;			
OC	Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE-BRAIN:			
RX	MEDLINE: 95292344.			
RA	Okada Y., Yamazaki H., Sekine-Aizawa Y., Hirokawa N.;			
RT	The neuron-specific kinesin superfamily protein KIF1A is a unique			
RT	monomeric motor for anterograde axonal transport of synaptic vesicle			
RT	precursors."			
RL	Cell 81:769-780(1995).			
RN	[2]			
RP	PRELIMINARY SEQUENCE OF 100-247 FROM N.A.			
RC	TISSUE-BRAIN:			
RC	MEDLINE: 93077686.			
RA	Aizawa H., Sekine Y., Takemura R., Zhang Z., Nangaku M.,			
RA	Hirokawa N.;			
RT	Kinesin family in murine central nervous system."			
RT	J. Cell Biol. 119:1267-1296(1992).			
CC	-1- FUNCTION: MOTOR FOR ANTEROGRADE AXONAL TRANSPORT OF SYNAPTIC			
CC	VESICLE PRECURSORS.			
CC	-1- SUBUNIT: MONOMER.			
CC	-1- TISSUE SPECIFICITY: EXPRESSED ALMOST EXCLUSIVELY IN ADULT BRAIN			
CC	TISSUE (MAINLY IN THE CEREBELLUM AND CEREBRUM) WITHIN A SINGLE			
CC	TYPE OF NEURONAL CELL. WITHIN THE NEURONAL CELL LEVELS ARE			
CC	CONCENTRATED AROUND THE AXON, WITH SMALLER AMOUNTS IN THE			
CC	PERINEURIAL AND SYNAPTIC REGIONS.			
CC	-1- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. UNC-104			
CC	SUBFAMILY.			
CC	-1- SIMILARITY: CONTAINS 1 FHA DOMAIN.			
CC	-1- SIMILARITY: CONTAINS 1 PH DOMAIN.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sdb.ch/announce/isb-sdb.ch).			
CC	or send an email to license@isb-sdb.ch .			
DR	EMBL: D29951; BAA06221.1; -.			
DR	PIR: E44259; E44259.			
DR	HSSP: P17119; 3KAR.			
DR	MGD: MGI:108391; KIF1A.			
DR	PRINTS: PRO0380; KINESINHEAVY.			
DR	PROSITE: PS00411; KINESIN MOTOR DOMAIN; 1.			
DR	PROSITE: PS50067; KINESIN MOTOR DOMAIN; 1.			
DR	PROSITE: PS50003; PH DOMAIN; 1.			
DR	PROSITE: PS50006; PH DOMAIN; 1.			
DR	PFAM: PF00169; PH; 1.			
DR	PFAM: PF00225; kinesin; 1.			
DR	PFAM: PF00498; FHA; 1.			
RW	Motor protein; Microtubules; ATP-binding; Coiled coil.			
FT	DOMAIN	1	361	MECHANOCHEMICAL (MOTOR).
FT	DOMAIN	366	383	COILED COIL (POTENTIAL).
FT	DOMAIN	429	462	COILED COIL (POTENTIAL).
FT	DOMAIN	516	572	COILED COIL (POTENTIAL).
FT	DOMAIN	622	681	COILED COIL (POTENTIAL).
FT	DOMAIN	801	822	COILED COIL (POTENTIAL).
FT	DOMAIN	1580</		

```

Query Match          10 38: Score 64; DB 1; Length 1695;
Best Local Similarity 23.48; Pred. No. 44;
Matches: 32; Conservative 19; Mismatches 56; Indels 30; Gaps
OY      PAPPV--CDSONECVGMWGV-----AYSAYVDVIRAAHEGYIEPEAR---GRID 52
        |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB      PAEPDPAFAQRELLEKGGIDMKOMEQRLOLEDQYRREREATYLLLEQQRLDYESKE 672
OY      53 ALREMIYYNTTENSKRDRRRRSY-----CHATWCFFRYDY-----VRSIWMHD- 101
        |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB      673 ALOKMDSRSTYPEVEVEEEDDEDEVOTERBCEALW-AFRKWIKYOFTSLDLDMGNAI 731
        |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY      101 ---TTNTISVSAHSVQ 114
        ||||
DB      732 FLKEANAISTVELAKRTVO 748

RESULT 10
MOG_MOUSE STANDARD: PRT; 246 AA.
ID MOG_MOUSE SEQUENCE OF 29-246 FROM N.A.
AC 061885; 062003; P70364;
DT 01-NOV-1997 (Rel. 35 Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE MYELIN-OLIGODENDROCYTE GLYCOPROTEIN PRECURSOR.
GN MOG.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
   Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
RP [1] SEQUENCE FROM N.A.
RC STRAIN=129;
RX MEDLINE; 95130110.
RA Daubas P., Pham-Dinh D., Dautigny A.;
RT "Structure and polymorphism of the mouse myelin/oligodendrocyte
   glycoprotein gene.";
RL Genomics 23:36-41(1994).
RN [2]
RP SEQUENCE FROM N.A.
RA Gardiner M.V., Mathieu J.M.;
RT "Murine and human MOG are highly conserved: cDNA analysis.";
RL Trans. Am. Soc. Neurochem. 24:234-234(1993).
RN [3]
RP SEQUENCE OF 29-246 FROM N.A.
RC STRAIN-BALB/C; TISSUE-BRAIN;
RX MEDLINE; 93376728.
RA Pham-Dinh D., Mattei M.-G., Nusbaum J.-L., Roussel G.,
   Pontorotti P., Roessel N., Mather I.H., Artzt K., Lindahl K.F.,
   Dautigny A.;
RT "Myelin/oligodendrocyte glycoprotein is a member of a subset of the
   immunoglobulin superfamily encoded within the major
   histocompatibility complex.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:7990-7994(1993).
RN [4]
RP SEQUENCE OF 29-54.
RC STRAIN-BALB/C; TISSUE-BRAIN;
RX MEDLINE; 92218912.
RA Ameliet P., Gardinier M.V., Zanetta J.-P., Mathieu J.-M.;
RT "Purification and partial structural and functional characterization
   of mouse myelin/oligodendrocyte glycoprotein.";
RL J. Neurochem. 58:1676-1682(1992).
CC -1- FUNCTION: MINOR COMPONENT OF THE MYELIN SHEATH. MAY BE INVOLVED IN
   COMPLETION AND/OR MAINTENANCE OF THE MYELIN SHEATH AND IN CELL-
   CELL COMMUNICATION.
CC -1- SUBUNIT: MAY FORM HOMODIMERS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: FOUND EXCLUSIVELY IN THE CNS, WHERE IT IS
   LOCALIZED ON THE SURFACE OF MYELIN AND OLIGODENDROCYTE CYTOPLASMIC
   MEMBRANES.
CC -1- DISEASE: REDUCED CONCENTRATIONS OF MOG ARE OBSERVED IN JIMPY AND
   QUACKING DYSPHAGNATING MUTANT MICE.
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
   ONE V-LIKE DOMAIN. BELONGS TO THE BTV/MOG SUBFAMILY.

```

CC - CAUTION: DO NOT CONFUSE MYELIN-OLIGODENDROCYTE GLYCOPROTEIN (MOG)
 CC WITH OLIGODENDROCYTE-MYELIN GLYCOPROTEIN (OMG).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)
 CC
 DR EMBL: L29503; AAC42023.1; JOINED.
 DR EMBL: L29498; AAC42023.1; JOINED.
 DR EMBL: L29500; AAC42023.1; JOINED.
 DR EMBL: L29501; AAC42023.1; JOINED.
 DR EMBL: L29499; AAC42023.1; JOINED.
 DR EMBL: L29502; AAC42023.1; JOINED.
 DR EMBL: U64572; AAB08096.1; -
 DR EMBL: L20942; AAA03180.1; -
 DR MGJ: MGJ:97435; MOG.
 DR PFAM: PF00047; 19; 1.
 DR Immunoglobulin domain; Glycoprotein; Myelin; Transmembrane; Signal.
 FT SIGNAL 1 28
 FT CHAIN 29 246
 FT DOMAIN 29 156
 FT TRANSMEM 157 177
 FT DOMAIN 178 209
 FT TRANSMEM 210 230
 FT DOMAIN 231 246
 FT TRANSMEM 45 133
 FT DOMAIN 52 126
 FT DISULFID 59 59
 FT CARBOHYD 21 21
 FT CONFLICT 21 21
 FT CONFLICT 32 32
 FT CONFLICT 95 95
 FT CONFLICT 169 169
 FT SEQUENCE 246 AA; 28271 MW; 1FLA84AD5CFB89 CRC64;
 Query Match 10.1%; Score 63; DB 1; Length 246;
 Best Local Similarity 37.5%; Pred. No. 7;
 Matches 15; Conservative 5; Mismatches 20; Indels 0; Gaps 0;
 QY 19 VGMGLAVASAVDVYRAAAHEGYIEPPARGRLDALREWI 58
 DB 65 VGMTRSPFSRVVHLYRNGKDDAQAQAPYRGRTELLKETI 104
 RESULT 11
 DAFI_STRCO STANDARD; PRT; 289 AA.
 AC 069695;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DE 15-DEC-1998 (Rel. 37, Last annotation update)
 DE DIAMINOPIMELATE EPIMERASE (EC 5.1.1.7) (DAP EPIMERASE).
 GN DAPF OR SC4H2.14.
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 NC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Seeger K.J., Harris D., Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: L1-2,6-DIAMINOHEPTANEDIOATE = MESO-
 CC DIAMINOHEPTANEDIOATE.
 CC -1- PATHWAY: SIXTH STEP IN THE BIOSYNTHESIS OF LYXINE FROM ASPARTATE
 CC SEMIALDEHYDE.
 CC -1- SUBCELLULAR LOCATION: CYTOSOL (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE DIAMINOPIMELATE EPIMERASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)
 CC
 DR EMBL: AL022268; CAAB330.1; -
 DR PROSITE: PS01326; DAP_EPIMERASE; 1.
 KW isomerase; Lysine biosynthesis.
 FT ACT_SITE 85 85
 FT ACT_SITE 218 218
 FT ACT_SITE 218 218
 FT SEQUENCE 289 AA; 30014 MW; 3F84933EFC61D18F CRC64;
 Query Match 10.1%; Score 63; DB 1; Length 289;
 Best Local Similarity 34.5%; Pred. No. 8.3;
 Matches 20; Conservative 8; Mismatches 12; Indels 18; Gaps 4;
 QY 5 ALPAPP-----VDSQNECVGMLGVASVVDYIRAAHEGYIEPPARGRLDALREWI 57
 DB 26 ALDPPAVALCDRR-----AGLEGDELHVVSAAH-----PEARQ-MAAEAEW 70
 RESULT 12
 CEGT_HUMAN
 ID CEGT_HUMAN STANDARD; PRT; 394 AA.
 AC Q16739;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE CERAMIDE GLUCOSYLTRANSFERASE (EC 2.4.1.80) (GLUCOSYLCERAMIDE SYNTHASE)
 DE (UDP-GLUCOSE:N-ACYLSPHINGOSINE D-GLUCOSYLTRANSFERASE) (GLCT-1)
 DE (FRAGMENT)
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
 NC Eutheria; Primates; Catarrhini; Homiidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 96209784.
 RA Ichikawa S., Sakiyama H., Suzuki G., Hidari K.I., Hirabayashi Y.;
 RT Expression cloning of a cDNA for human ceramide glucosyltransferase
 RT that catalyzes the first glycosylation step of glucosylphingolipid
 RT synthesis.
 RL Proc. Natl. Acad. Sci. U.S.A. 93:4638-4643(1996).
 CC -1- FUNCTION: MAY SERVE AS "FLIPASE" AS WELL AS A GLUCOSYLTRANSFERASE
 CC THAT TRANSFERS GLUCOSE TO CERAMIDE.
 CC -1- CATALYTIC ACTIVITY: UDP-GLUCOSE + N-ACYLSPHINGOSINE =
 CC UDP + D-GLUCOSYL-N-ACYLSPHINGOSINE.
 CC -1- PATHWAY: FIRST GLYCOSYLATION STEP OF GLUCOSYLINOLIPID SYNTHESIS.
 CC -1- SUBCELLULAR LOCATION: TYPE III MEMBRANE PROTEIN. ENDOPLASMIC
 CC RETICULUM.
 CC -1- TISSUE SPECIFICITY: FOUND IN ALL TISSUES EXAMINED.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)
 CC
 DR EMBL: D50840; BAA09451.1; -
 DR Transferrase; Glycosyltransferase; Transmembrane; Signal-anchor;
 KW Endoplasmic reticulum.
 FT DOMAIN 1 10
 FT TRANSMEM 11 31
 FT TRANSMEM 286 306
 FT TRANSMEM 314 334
 FT NON_TER 394 394
 FT SEQUENCE 394 AA; 44853 MW; 3B998569F8A96449 CRC64;
 LUMENAL (POTENTIAL).
 SIGNAL-ANCHOR (TYPE-IB MEMBRANE PROTEIN)
 (POTENTIAL).
 POTENTIAL.
 POTENTIAL.

FT CONFLICT 589 589 Y -> C (IN REF. 3).
 FT CONFLICT 601 601 V -> A (IN REF. 3).
 FT CONFLICT 1364 1364 E -> A (IN REF. 3).
 FT CONFLICT 2118 2119 ML -> IV (IN REF. 1).
 SQ SEQUENCE 4092 AA; 471337 MW; 309DF447E8E2D6B8 CRC64;

Query Match 10.1%; Score 63; DB 1; Length 4092;
 Best Local Similarity 27.3%; Pred. No. 1.5e+02;
 Matches 18; Conservative 13; Mismatches 33; Indels 2; Gaps 2;

QY 47 ARGDALRWITTYNTTSSKRRRRSVCHARTWCFCKXYRRSIMDITTTTIS 106
 DB 3667 ARTRVDEIMLWLYOEYVCPSTALDKKFKMIM-AMTWELCYKFD-IESEQYKAVLTMIG 3724
 QY 107 VVSAS 112
 DB 3725 VLSSE 3730

RESULT 15

MOG_RAT STANDARD: PRT; 245 AA.

ID MOG_RAT

AC 063345;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE MYELIN-OLIGODENDROCYTE GLYCOPROTEIN PRECURSOR.

GN MOG.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

RN [1]

RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

RX MEDLINE; 93085763.

RA Gardiner M.V., Amiguet P., Linington C., Mathieu J.-M.;

RT "Myelin/oligodendrocyte glycoprotein is a unique member of the

immunoglobulin superfamily."

RL J. Neurosci. Res. 33:177-187(1992).

RN [12]

RP SEQUENCE OF 28-245 FROM N.A.

RX TISSUE-BRAIN;

RA MEDLINE; 93376728.

RA Pham-Dinh D., Mettel M.-G., Nussbaum J.-L., Rousset G.,

Pontarotti P., Roessel N., Mather I.H., Artzt K., Lindahl K.F.,

Dautigny A.;

RT Myelin/oligodendrocyte glycoprotein is a member of the

immunoglobulin superfamily encoded within the major

histocompatibility complex."

RL Proc. Natl. Acad. Sci. U.S.A. 90:7990-7994(1993).

RN [13]

RP STRUCTURE BY NMR OF 62-82.

RX MEDLINE; 97354172.

RA Albouze-Abou S., Wilson J.C., Bernard C.C.A., von Itzstein M.;

RT "A conformational study of the human and rat encephalitogenic myelin

oligodendrocyte glycoprotein peptides 35-55."

RL Eur. J. Biochem. 246:59-70(1997).

RN [1]

RP FUNCTION: MINOR COMPONENT OF THE MYELIN SHEATH. MAY BE INVOLVED IN

COMPLETION AND/OR MAINTENANCE OF THE MYELIN SHEATH AND IN CELL-

CELL COMMUNICATION.

-1 SUBUNIT: MAY FORM HOMODIMERS.

-1 SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.

-1 TISSUE SPECIFICITY: FOUND EXCLUSIVELY IN THE CNS, WHERE IT IS

LOCALIZED ON THE SURFACE OF MYELIN AND OLIGODENDROCYTE CYTOPLASMIC

MEMBRANES.

-1 DEVELOPMENTAL STAGE: A PEAK OF EXPRESSION HAS BEEN OBSERVED

BETWEEN POSTNATAL DAYS 15 AND 25, COINCIDENT WITH THE PERIOD OF

ACTIVE MYELINATION.

-1 SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS

ONE V-LIKE DOMAIN. BELONGS TO THE BTN/MOG SUBFAMILY.

-1 CAUTION: DO NOT CONFUSE MYELIN-OLIGODENDROCYTE GLYCOPROTEIN (MOG)

WITH OLIGODENDROCYTE-MYELIN GLYCOPROTEIN (OMG).

CC

CC

CC

CC

CC

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; M99485; AAA41628.1; -

DR EMBL; L21995; -; NOT_ANNOTATED_CDS.

DR PFM; PF00047; 19; 1.

KW Immunoglobulin domain; Glycoprotein; Myelin; Transmembrane; Signal.

FT SIGNAL

FT CHAIN

FT DOMAIN

FT TRANSMEM

FT DOMAIN

FT TRANSMEM

FT DOMAIN

FT DOMAIN

FT DISULFD

FT CARBOHYD

SQ SEQUENCE

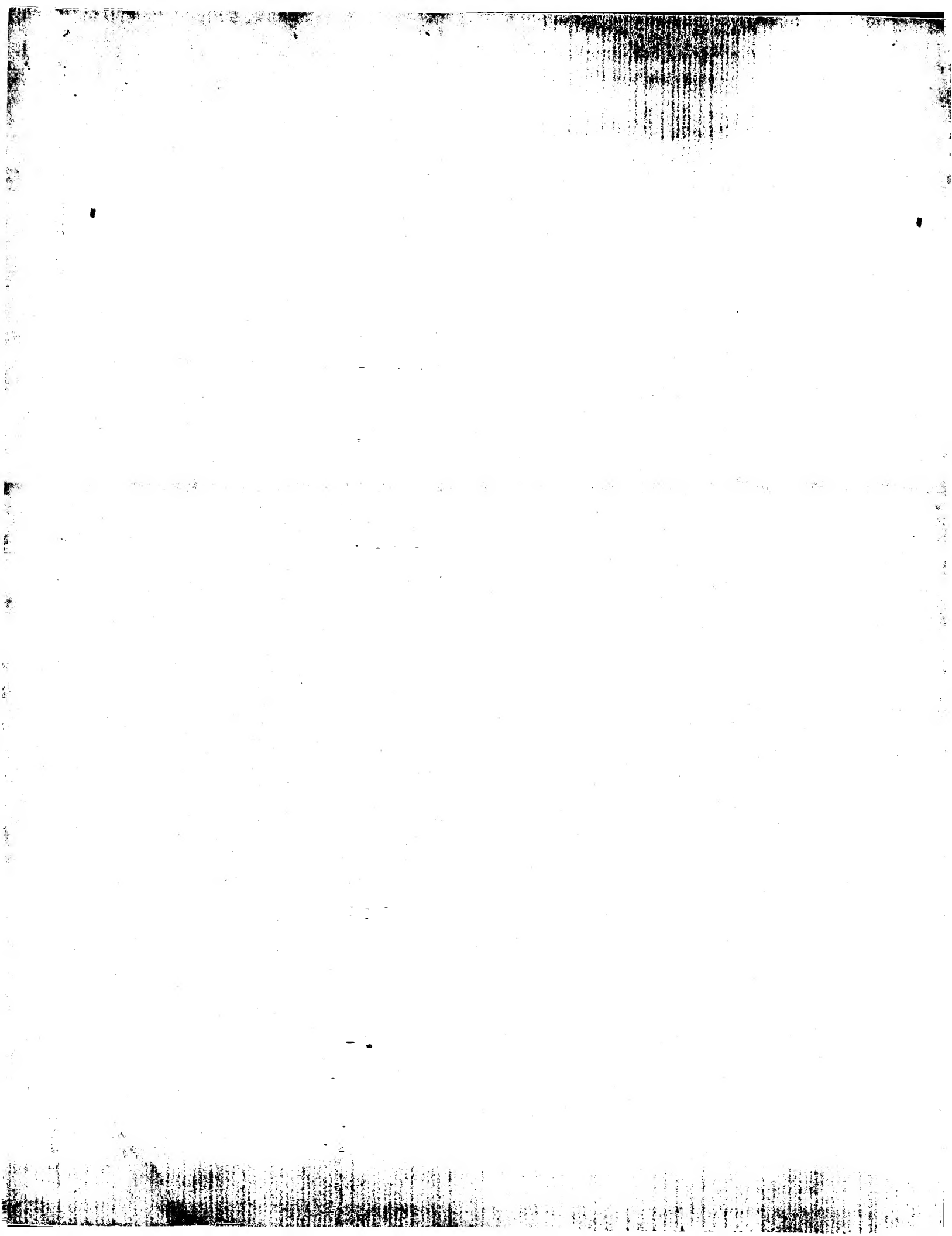
Query Match 10.0%; Score 62; DB 1; Length 245;
 Best Local Similarity 37.3%; Pred. No. 9;
 Matches 15; Conservative 5; Mismatches 20; Indels 0; Gaps 0;
 QY 19 VGMGVAYSAYVDYVRAAHGVTIEPEARGRLALREWI 58
 DB 64 VGMVRSPEPSRVVHLIRNGKDDAQAQAEYRRTLESTI 103

Search completed: June 23, 2000, 10:14:51
 Job time: 672 sec

Mon Jun 26 09:11:01 2000

us-09-214-478-4.rsp

Page 9



OM protein - protein search, using sw model

Run on: June 23, 2000, 10:07:44 ; Search time 61.44 Seconds
(without a) {comments}

128.647 Million cell updates/sec

```

Title: US-09-214-478-4
Perfect score: 621
Sequence: 1 MWLPALPAPEVCDQNECVG.....SIWHDTTNTISVSAHSVQ 114

```

Scoring table:

BLOSUM62	Gapenx 0.5
Gapen 10.0	

Searched: 225878 seqs, 69334122 residues

Total number of hits satisfying chosen parameters: 225878

Maximum DB seq length: 100000000

Post-processing: Minimum Match 08
Listing First 45 summaries

Database

```

1: SP_RfamBL_12.*
2: 1:  sp_archaea:*
3: 2:  sp_bacteria:*
4: 3:  sp_fungi:*
5: 4:  sp_human:*
6: 5:  sp_invertebrate:*
7: 6:  sp_mammal:*
8: 7:  sp_mhc:*
9: 8:  sp_organelle:*
10: 9:  sp_phase:*
11: 10: sp_plant:*
12: 11: sp_rodent:*
13: 12: sp_virus:*
14: 13: sp_vertebrate:*
15: 14: sp_unclassified:

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	Match	Length	DB	ID	Description
No.	Score					
1	254.5	41.0	120	12	P89082	P89082 human adeno
2	183	29.5	121	12	Q64866	Q64866 human adeno
3	75	12.1	347	5	Q25436	Q25446 mayefiola d
4	72.5	11.7	1232	1	Q58318	Q58318 methanococ
5	66.5	10.7	945	10	Q94VK3	Q94VK3 arabidopsis
6	66	10.6	559	12	Q05552	Q05552 ectromelia
7	65	10.5	527	11	Q64304	Q64304 rattus leuc
8	64.5	10.4	284	2	Q53718	Q53718 mycobacteri
9	64.5	10.4	496	10	Q92Q98	Q92998 arabidopsis
10	64.5	10.4	1095	13	Q93596	Q93596 brachydanio
11	64	10.3	1083	10	Q48839	Q48839 arabidopsis
12	63.5	10.2	264	5	Q19843	Q19843 caenorhabdi
13	63.5	10.2	386	13	Q93370	Q93370 brachydanio
14	63.5	10.2	516	3	Q94512	Q94512 schizosach
15	63.5	10.2	756	10	Q22863	Q22863 arabidopsis
16	63.5	10.2	1624	2	Q53203	Q53203 mycobacteri
17	63.5	10.2	1639	2	Q85019	Q85019 mycobacteri
18	63	10.1	394	11	Q55149	Q55149 rattus norvi
19	63	10.1	394	11	Q88693	Q88693 mus musculu
20	63	10.1	896	2	Q50747	Q50747 borrelia bu

21	62.5	10.1	239	11	035558'
22	62.5	10.1	240	11	054832
23	62.5	10.1	322	2	086892
24	62.5	10.1	2843	4	09Y6K7
25	62	10.0	211	3	012346
26	62	10.0	629	2	09ZNP9
27	62	10.0	895	1	058052
28	62	10.0	969	5	027502
29	62	10.0	1759	5	093187
30	62	10.0	1781	5	017758
31	61.5	9.9	244	5	018352
32	61.5	9.9	455	10	09ZMS2
33	61.5	9.9	776	10	023943
34	61.5	9.9	886	4	09Y599
35	61.5	9.9	1413	4	075656
36	61.5	9.9	1661	5	077244
37	61.5	9.9	2080	4	075923
38	61	9.8	84	6	019086
39	61	9.8	215	12	080635
40	61	9.8	308	2	02X9R1
41	61	9.8	668	12	056765
42	60.5	9.7	210	2	P95221
43	60.5	9.7	223	2	050352
44	60.5	9.7	898	6	062683
45	60	9.7	151	2	007408

ALIGNMENTS

RESULT	1		
P89082		PRELIMINARY;	PRT; 120 AA.
AC	P89082;		
DI	01-MAY-1997 (TRIMBLrel_03, Created)		
DT	01-May-1997 (TRIMBLrel_03, Last sequence update)		
DT	01-NOV-1998 (TRIMBLrel_08, Last annotation update)		
DE	ORF4.		
OS	Human adenovirus type 9.		
OC	Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE; 96327849.		
RA	JAVIER R.; SHENK T.;		
RT	"Mammary tumors induced by human adenovirus type 9: a role for the		
RI	viral early region 4 gene.";		
RL	Breast Cancer Res. Treat. 39:57-67(1996).		
DR	EMBL; S82508; AAB37507.1; "		
SQ	SEQUENCE 120 AA; 14152 MW; EE2C2416 CRC32;		

Query Match	41.0%;	Score 254.5;	DB 12;	Length 120;
Best Local Similarity	40.7%;	Pred. No. 5.6e-21;		
Matches 46;	Conservative 26;	Mismatches 40;	Indels 1;	Gaps 1.

QY	1	MFLPALPAPVCDSONECVGWLGVAIVSAVDVIIRAAHEGYIIEEPARGRLDALREWIY	60
Dd	1	: : : : : : : : : : : : : : : : : : : :	59
		1 MVLPLPPPLDLNDROGS - INMGMAVRVLADVMRGIRMDGFSSIAAEELLNLTREMYF	
QY	61	NYTTESKRRDARRRSVCARTWCECRKXDYRSLWDHTTTNTISVSANSHV	113
Dd	60	:: :: ::: : : : : :: : : : : : : : : :	112
		60 SWMTERQQQRKGRCGRGICCSRAHFMCWKDYKRRIHYNANDSIOLAPPSTI	
RESULT	2		
Q64866	ID	Q64866	PRELIMINARY; PRT; 121 AA.
AC	Q64866;		
DT	01-NOV-1996	(TREMBLrel. 01, Created)	
DT	01-NOV-1996	(TREMBLrel. 01, Last sequence update)	
DE	01-NOV-1998	(TREMBLrel. 08, last annotation update)	
OS	ADENOVIRUS TYPE 40,	COMPLETE GENOME.	
	Human adenovirus type 40.		

OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DUGAN;
 RX MEDLINE: 87122182.
 RA IISHINO M., SAMADA Y., YABGASHI T., DEMURA M., FUJINAGA K.,
 RT "Nucleotide sequence of the adenovirus type 40 inverted terminal
 RL repeat: close relation to that of adenovirus type 5.";
 RN Virology 156:414-416(1987).
 RP SEQUENCE FROM N.A.
 RC STRAIN-DUGAN;
 RX MEDLINE: 88084437.
 RA VAN LOON A.E., LIGTENBERG M., REEMST A.M., SUSSENBACH J.S.,
 RT "Structure and organization of the left-terminal DNA regions of
 RL fastidious adenovirus types 40 and 41.";
 RN Gene 58:109-126(1987).
 RP SEQUENCE FROM N.A.
 RC STRAIN-DUGAN;
 RX IISHINO M.;
 RT Sapporo Igaku Zasshi 57:59-66(1988).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DUGAN;
 RX MEDLINE: 88265890.
 RA IISHINO M., OHASHI Y., EMOTO T., SAWADA Y., FUJINAGA K.,
 RT "Characterization of adenovirus type 40 EL region.";
 RL Virology 165:95-102(1988).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DUGAN;
 RX MEDLINE: 88160034.
 RA VOS H.L., VAN DER LEE F.M., REEMST A.M., VAN LOON A.E.,
 RT "The genes encoding the DNA binding protein and the 23k protease of
 RL adenovirus types 40 and 41.";
 RN Virology 163:1-10(1988).
 RP SEQUENCE FROM N.A.
 RC STRAIN-DUGAN;
 RX MEDLINE: 89370295.
 RA KIDD A.H., ERASMUS M.J.;
 RT "Sequence characterization of the adenovirus 40 fiber gene.";
 RL Virology 172:134-144(1989).
 RN [7]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DUGAN;
 RX MEDLINE: 90111698.
 RA TOOGOOD C.I., MURALI R., BURNETT R.M., HAY R.T.;
 RT "The adenovirus type 40 hexon: sequence, predicted structure and
 RL relationship to other adenovirus hexons.";
 RN J. Gen. Virol. 70:3203-3214(1989).
 RP SEQUENCE FROM N.A.
 RC STRAIN-DUGAN;
 RX MEDLINE: 94087748.
 RA DAVISON A.J., TELFORD E.A., WATSON M.S., MCBRIDE K., MAUTNER V.;
 RT "The DNA sequence of adenovirus type 40.";
 RL J. Mol. Biol. 234:1308-1316(1993).
 RN [9]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DUGAN;
 RX PIENTIAER N.J., SLEEMENDA S.B., PIENTIAER D., LUFFTIG R.B.;
 RT Submitted (JUN-1993) to the EMBL/GenBank/DBJ databases.
 RL EMBL: L19443; A013982.1;
 DR SEQUENCE 121 AA; 13976 MW; 8884054F CRC32;

Query Match 29.5%; Score 183; DB 12; Length 121;
 Best Local Similarity 35.5%; Pred. No. 4.6e-13;
 Matches 38; Conservative 20; Mismatches 49; Indels 0; Gaps 0;

OC 1 MYLPAPPCVDSQNECVGLGVAYSANVDYIRAAHEGVTIEEPARGDLALREWY 60
 DB 1 MFLPSLPPEPVSDGLCLSWLELALTCCLNVDYIRYNSISPRBELLSGLQEWLV 60
 QY 61 NYTTERSKRRDRRRRSVCARTWFCRKYDYRSTRIMHDTNTSY 107
 DB 61 ALKTEHARQRNRFRANICQRYRFLKQTEAVRELLITDATQWYSL 107

RESULT 3
 ID 025436 PRELIMINARY; PRT; 347 AA.
 AC 025436;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DT 01-MAY-1999 (TREMblrel. 10, Last annotation update)
 DE TRANSPOSASE.
 OS Mayetiola destructor (Hessian fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Sciarioidea;
 OC Cecidomyiidae; Mayetiola.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TRANSPOSON-DESMARL;
 RA RUSSELL V.W., SHUKLE R.H.;
 RT Submitted (Apr-1995) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U24436; AA66077.1;
 DR PFM: PF01359; Transposase.1;
 SQ SEQUENCE 347 AA; 40758 MW; AEF5782 CRC32;

Query Match 12.1%; Score 75; DB 5; Length 347;
 Best Local Similarity 26.3%; Pred. No. 1.4;
 Matches 26; Conservative 16; Mismatches 49; Indels 8; Gaps 3;

QY 3 LPALPAPPCVDSQNECVGLGVAYSANVDYIRAAH--EGVTIEEPARGDLALREWY 59
 DB 79 LEALDEDCOCQGEELANSGLVTOQALSKRLKAGYIOKQGNWPHLEKPR-DVERRRCM 137
 QY 60 NYTTERSKRRDRRRRSVCARTWFCF---RKDYR 94
 DB 138 SEMLLQRHKKRFSRLITGDEKWIHYDNRKRSYKR 176

RESULT 4
 ID 058318 PRELIMINARY; PRT; 1232 AA.
 AC 058318;
 DT 01-JAN-1998 (TREMblrel. 05, Created)
 DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)
 DT 01-JAN-1998 (TREMblrel. 05, Last annotation update)
 DE HYPOTHEICAL PROTEIN M0908.
 GN M0908.
 OS Methanococcus jannaschii.
 OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
 OC Methanococcus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 9637999;
 RA BUTT C.J., WHITE O., OLSEN G.J., ZHOU L., FLEISCHMANN R.D.,
 RA BUTTON G.G., BLAKE J.A., FITZGERALD L.M., CLAYTON R.A., GOCAYNE J.D.,
 RA KERLAVAGE A.R., DOUGHERTY B.A., TOMB J.-F., ADAMS M.D., REICH C.I.,
 RA OVERBERG R., KIRKNESS E.F., WEINSTOCK K.G., MERRICK J.M., GLODER A.,
 RA SCOTT J.L., GEORHAGEN N.S.M., WEIDMAN J.F., FUHRMAN J.L., NGUYEN D.,
 RA UTTERBACK T.R., KELLEY J.M., PETERSON J.D., SADOW P.W., HANNA M.C.,
 RA COTTON M.D., ROBERTS K.M., HURST M.A., KATNE B.P., BORDOVSKI M.,
 RA KLENK H.-P., FRASER C.M., SMITH H.O., WOESE C.R., VENTER J.C.;
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
 jannaschii.";
 RT Science 273:1058-1073(1996).
 CC -1- SIMILARITY: STRONG TO P.DENTRIFICANS COBN AND M.JANNASCHII
 M01441.

DR EMBL: U67534; AAB98910.1; -
RC TYGR: M07098; -
KW Hypothetical protein.
SQ SEQUENCE 1232 AA; 142643 MW; FB712545 CRC32;

Query Match 11.7%; Score 72.5; DB 1; Length 1232;
Best Local Similarity 23.8%; Pred. No. 12;
Matches 35; Conservative 14; Mismatches 51; Indels 47; Gaps 6;

QY 3 LPALPAPVPC-----DSQECVGLVAISAV-----30
DB 86 IPEPLPTISIHDPLEKEDRVDDVKNKYKLG--YEGVYNNKLLLYLANFGNLNV 143
OY 30 VDVYRAAHAGVYIEPARRLALREMYNYNTERSKRRDRRSYCHAFMPCFRK 88
DB 144 EYEPBPMPWGGIYKGYFETLDD---LYNLKELGRDLKPIIGVLFYRMWFVANN 198
OY 89 YDYVRSIWHDTTN---TISVSAH 111
DB 199 IDYVNDLI--DIENKGAIPAVSSH 223

RESULT 5
092VK3 PRELIMINARY; PRT; 945 AA.

AC 092VK3;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DE 01-MAY-1999 (TREMBLrel. 10, Last annotation update)
DE REVERSE-TRANSCRIPTASE-LIKE PROTEIN.
GN T6B13.12.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA;
RA ROUNSLEY S.D., LIN X., KAUL S., SHEA T.P., FUJII C.Y., MASON T.M.,
RA SHEN M., RONNING C.M., FRASER C.M., SOMERVILLE C.R., VENTER J.C.;
RA "Arabidopsis thaliana chromosome II BAC T6B13 genomic sequence."
RL Submitted (OCT-1998) to the EMBL/Genbank/DBJ databases.
DR EMBL: AC005398; AAC69378.1; -
SQ SEQUENCE 945 AA; 108084 MW; 88791777 CRC32;

Query Match 10.7%; Score 66.5; DB 10; Length 945;
Best Local Similarity 20.7%; Pred. No. 40;
Matches 23; Conservative 16; Mismatches 29; Indels 43; Gaps 6;

QY 15 QNECGWGLVAVSAVDYVRAAAHEGV-----IPEERGRDLNR-EMYYNYT 64
DB 180 QNKYVGM-----TQEVLVGAFMNGLYSINGIRMPQPSRLREYINLRVGM----- 226
OY 65 ESKRRDRRRRSYCH-----ARTWFCFRXY-----DYVRSIM 97
DB 226 -KGRSDRRKRAITHLHLAVLSITWCAKROSNRRORDSVKKNYERKGVW 275

RESULT 6
005552 PRELIMINARY; PRT; 559 AA.

AC 005552;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE PROTEIN P65 (PROTEIN B21 HOMOLOG).
OS Ectromella virus.
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Orthopoxvirus.
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN-MILL HILL;
RA MEDLINE: 94091057.
RX SENKEVICH T.G., MURAVNIK G.L., POZDNYAKOV S.G., CHIZHIKOV V.E.,
RA KIZANKINA O.I., SHCHELKUNOV S.N., KOONIN E.V., CHERNOS V.I.;
RT "Nucleotide sequence of XhoI O fragment of ectromella virus DNA
RT reveals significant differences from vaccinia virus."
RL Virus Res. 30:73-88(1993).
DR EMBL: Z14256; CAA78622.1; -
DR PRAM: PF00651; B7B: 1.
DR PRAM: PF01344; Ketch: 3.
SQ SEQUENCE 559 AA; 65281 MW; 491EAE41 CRC32;

Query Match 10.6%; Score 66; DB 12; Length 559;
Best Local Similarity 20.3%; Pred. No. 25;
Matches 26; Conservative 26; Mismatches 42; Indels 34; Gaps 7;

QY 1 MVLPAFPVCDSONECVGLVAVSAVDYVRAAAHEGVYIEPARGRDLALREMYNY 60
DB 150 MYMKNIPTLTITDAFKNAV-----FEILDII--STNDEYVYRGY-KVTILKWL DY 200
OY 61 NYTERS-----KRRDRRRRSYCHARTW-----CFR-----KYDYVRS--- 96
DB 201 NYTEEQLCLTSCIDQINDKRSRLLYSNTITMTSSCVKFLDNKKNNIIPQLCL 260
QY 96 IWHDTTN 103
DB 261 VYHDTN 268

RESULT 7
064304 PRELIMINARY; PRT; 527 AA.

AC 064304;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE PROTEIN LINKED TO SYSTEM L-LIKE NEUTRAL AMINO ACID TRANSPORT
DE ACTIVITY.
GN LAT.
OS Rattus leucopus (mottle-tailed rat), and Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE: 96128073.
RA BROEER S., BROEER A., HAMPRECHT B.;
RT "The 4F2c surface antigen is necessary for expression of system L-
RT like neutral amino acid-transport activity in C6-BU-1 rat glioma
RT cells: evidence from expression studies in Xenopus laevis oocytes."
RL Biochem. J. 312:863-870(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE DAWLEY; TISSUE-JETUNAL BRUSH BORDER;
RA YAO S.Y.M., MOZYKA W.R., CHEESEMAN C.I., ELLIOTT J.F., YOUNG J.D.;
RL Submitted (MAY-1996) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-SPLAGUE-DAWLEY; TISSUE-LIVER;
RX MEDLINE: 98395066.
RA KANAI Y., SEGAWA H., MIYAMOTO K., UCHINO H., TAKEDA E., ENDOU H.;
RT "Expression cloning and characterization of a transporter for large
RT neutral amino acids activated by the heavy chain of 4F2 antigen
RT (CD98)."
RL J. Biol. Chem. 273:23629-23632(1998).
DR EMBL: X89325; CAA61509.1; -
DR EMBL: U59324; AAC53560.1; -
DR EMBL: AB015433; BAA33036.1; -
DR PRAM: PF00128; alpha3-amyase; 1.
SQ SEQUENCE 527 AA; 58072 MW; 6C2869F1 CRC32;

Query Match 10.5%; Score 65; DB 11; Length 527;
Best Local Similarity 29.5%; Pred. No. 30;
Matches 18; Conservative 10; Mismatches 13; Indels 20; Gaps 3;

OC 19 VGMGVASVAVDVIRAA-----HEG-----VTEPARGRLALREMY 59
DB 84 LGWGMAGAVVITVAPRCRELFPVQRMWKGALYRIGDLQAFVGPBARG-IGLKNHLE 142
DY 60 Y 60
DB 143 Y 143

RESULT 8
ID 053718; PRELIMINARY; PRT; 284 AA.

OC 01-JUN-1998 (TRENBLREL. 06, Created)
DT 01-JUN-1998 (TRENBLREL. 06, Last sequence update)
DE 01-NOV-1998 (TRENBLREL. 08, Last annotation update)
GN HYPOHETICAL 31.8 KD PROTEIN.

OC Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RA BROWN D., CHURCHER C.M.;
RL Submitted (FEB-1998) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RA COLE S.T., PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;
RL Submitted (FEB-1997) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RA MEDLINE; 96181548.
RA PHILIPP W.J., POULET S., EIGLMEIER K., PASCOPELLA L.,
RA BALASUBRAMANIAN V., HEYM B., BERGH S., BLOOM B.R., JACOBS W.R. JR.,
RA COLE S.T.;
RT "An integrated map of the genome of the tubercle bacillus,
RT Mycobacterium tuberculosis H37RV, and comparison with Mycobacterium
RT leprae.";
RL Proc. Natl. Acad. Sci. U.S.A. 93:3132-3137(1996).
DR EMBL; AL021931; CNA17389.1;
KW Hypothetical protein.
SQ SEQUENCE 284 AA; 31801 MW; 7A6BB7C1 CRC32;

Query Match 10.4%; Score 64.5; DB 2; Length 284;
Best Local Similarity 30.4%; Pred. No. 16;
Matches 17; Conservative 9; Mismatches 21; Indels 9; Gaps 2;

OC 63 YTERSKRRDRRRRSVCHARTWCFKRYDVRSI-----WHTTNTITSVSAHSV 113
DB 21 YVDIDRRKRSKR-----KSWARSHGFDERSTELIKRWTRGVASTGVAAHNV 72

RESULT 9
ID 092098; PRELIMINARY; PRT; 496 AA.

OC 01-MAY-1999 (TRENBLREL. 10, Created)
DT 01-MAY-1999 (TRENBLREL. 10, Last sequence update)
DE 01-NOV-1999 (TRENBLREL. 12, Last annotation update)
GN PUTATIVE GLUCOSYL TRANSFERASE.
RA Arabidopsis thaliana (Mouse-ear cress).
OC Eudicotyledons; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;

OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA;
RA LIN X., KAU S., SHEA T.P., FUJII C.Y., SHEN M., VANAKEN S.E.,
RA BARNSTEAD M.E., MASON T.M., BOWMAN C.L., RONNING C.M., BENITO M.,
RA CARBERA A.J., CREASY T.H., BUELL C.R., TOWN C.D., NIERMAN W.C.,
RA FRASER C.M., VENTER J.C.;
RT "Arabidopsis thaliana chromosome II BAC F13K3 genomic sequence.";
RL Submitted (MAR-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL; AC006282; AAD20152.1;
DR PROSITE; PS00375; UDPGT. 1.
KW Transferase.
SQ SEQUENCE 496 AA; 55991 MW; 2D7E8306 CRC32;

Query Match 10.4%; Score 64.5; DB 10; Length 496;
Best Local Similarity 26.7%; Pred. No. 31;
Matches 20; Conservative 12; Mismatches 24; Indels 19; Gaps 3;

OC 15 QNECVGWLG-----VAISAVDVIRAAHE-GVTEPEAR-----GRDLALR 55
DB 275 QDECIKWLDSKDVESLYVCLGSLCNPLAQLRELDGLLEATKRFPIWIRGGKXHELA 334
DY 56 EWYYNYTERSKRR 70
DB 335 EWILSGFEREKER 349

RESULT 10
ID 093596; PRELIMINARY; PRT; 1095 AA.

OC 01-NOV-1998 (TRENBLREL. 08, Created)
DT 01-NOV-1998 (TRENBLREL. 08, Last sequence update)
DE 01-NOV-1999 (TRENBLREL. 12, Last annotation update)
GN PROTEIN TYROSINE KINASE (EC 2.7.1.112).
RA JAK2A.
OC Brachydanio rerio (Zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
OC Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Cyprininae; Rasbora; Danio.
RN [1]
RP SEQUENCE FROM N.A.
RA OATES A.C.;
RL Thesis (1998), University of Melbourne, Australia.
DR EMBL; AL005690; CNA06674.1;
DR HSPD; P08631; ZHCK.
DR PROSITE; PS00107; PROTEIN_KINASE_AAP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PFMW; PF00069; PKINASE; 2.
KW Transferase.
SQ SEQUENCE 1095 AA; 125171 MW; 961AB4C1 CRC32;

Query Match 10.4%; Score 64.5; DB 13; Length 1095;
Best Local Similarity 22.4%; Pred. No. 79;
Matches 19; Conservative 17; Mismatches 36; Indels 13; Gaps 2;

OC 5 ALPAPVCDSONECVGLGVAISAVDVIRAAHEGVTEPEARGLDALREMYNYTN 64
DB 142 AVSPVSLFEQECIG-----LAVLIDMLRAKEG-----KSPVDIYHSSYSFLP 188

OC 65 ERSKRRDRRRSVCHARTWCFKRY 89
DB 189 KMMREHIONOHVTRKIRFRFRF 213

RESULT 11
ID 048839; PRELIMINARY; PRT; 1083 AA.

ID	AC	DT	DE	GN	OS
SO	EMBL: U49829; AAA93385.1; "				
DR	SEQUENCE	264 AA; 29808 MW; 90319321 CRC32;			
Query Match					
Best Local Similarity 23.2%; Pred. No. 19;					
Matches 26; Conservative 17; Mismatches 48; Indels 21; Gaps 6;					
OY	13	DSQECV---GW---LGVAYSADVDIRAAHEGYIEPEARGRLDALEMIYNYTT	64		
DB	6	DGHIVCVCSINGSGEMTGLSTCYTSKFAVNGAM-ESLQMEHLDRG-LEGITTTLLPYFA	63		
OY	65	ERSKRDRBRRSVCHARTWECF-----RKYYVRSIWHDTTINISVY	108		
DB	64	RTPILENNRPTC--TWPEMSIRSCSKHMDSLKEKVAHVPSITTL	112		
RESULT 13					
ID	093370	PRELIMINARY;	PRT;	386 AA.	
AC	093370;				
DT	01-NOV-1998 (TREMBLrel, 08, Created)				
DT	01-NOV-1998 (TREMBLrel, 08, Last sequence update)				
DT	01-NOV-1999 (TREMBLrel, 12, Last annotation update)				
DE	TRANSCRIPTION FACTOR.				
GN	PAX2.2.				
OS	Brachydanio rerio (zebrafish) (zebra danio).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;				
OC	Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;				
OC	Cyprinoidae; Cyprinidae; Rasbora; Danio.				
RC	[1]				
RP	SEQUENCE FROM N.A.				
RA	MEDLINE; 98337826.				
RX	PEPPER P.L.; GERSTER T., LUN K., BRAND M., BUSSLINGER M.;				
RT	"Characterization of three novel members of the zebrafish Pax2/5/8				
RT	family: dependency of Pax5 and Pax8 expression on the Pax2.1 (nol)				
RT	function.";				
RL	Development 125:3063-3074(1998).				
CC	-1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).				
CC	-1- SIMILARITY: CONTAINS A PAIRED BOX DOMAIN.				
DR	EMBL: AF072547; AAC31811.1; "				
DR	HSP: P06601; IPDN.				
DR	PROSITE; PS00034; PAIRED_BOX; 1.				
DR	PFAM; PF00292; PAX; 1.				
DR	PRINTS; PR00027. PAIRED_BOX.				
KW	Paired box; DNA-binding; Developmental protein; Nuclear protein.				
SO	SEQUENCE	386 AA; 41541 MW; 5BCDF72B CRC32;			
Query Match					
Best Local Similarity 10.2%; Score 63.5; DB 13; Length 386;					
Matches 16; Conservative 22; Mismatches 25; Indels 11; Gaps 2;					
OY	2	VLPALPAPVCDSONECGMLGVAYSANVYDRI-----AAHSEVYIEPPARGRLD	52		
DB	162	IIFSPASPEVSSSSNDPYG--SYINGILGPSNGEKRRKRDADGSEGSASDSQGSVE	219		
OY	53	ALREMIYNYTTER	66		
DB	220	SLRKHRLADATQO	233		
RESULT 14					
ID	094512	PRELIMINARY;	PRT;	516 AA.	
AC	094512;				
DT	01-MAY-1999 (TREMBLrel, 10, Created)				
DT	01-MAY-1999 (TREMBLrel, 10, Last sequence update)				
DT	01-MAY-1999 (TREMBLrel, 10, Last annotation update)				
DE	PROBABLE INVOLVEMENT IN ERGOSTEROL BIOSYNTHESIS.				
GN	SPR646.08C.				
OS	Schizosaccharomyces pombe (Fission yeast).				
OC	Eukaryota; Fungi; Ascomycota; Archia; Saccomycetes;				

OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-972H-;
 RA SEEGER K., HARRIS D., WOOD V., RAJANDREAM M.A., BARRELL B.G.;
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL035216; CAA22812.1; -
 SO SEQUENCE 516 AA; 58692 MW; BA471658 CRC32;

Query Match 10.2%; Score 63.5; DB 3; Length 516;
 Best Local Similarity 22.9%; Pred. No. 42;
 Matches 25; Conservative 14; Mismatches 47; Indels 23; Gaps 4;

QY 21 WGVAVASAV-----VDVRAAHGEGVTEPEARGRLDRLREWIYNYTTERS 67
 DB 354 WIGKPSLVEGYCYRDPNDTDSIKAVPKENILATFGN-----WRNCIFYSGESE 408
 QY 68 KRR--DRRRSYCHARTWFCFRKYDVRRSIWHDITNTISVSAHSVQ 114
 DB 409 SRMLVDLNLDELVHKRCPLDKQFPESRKIMFPVTHN--ILAKHYTQ 454

RESULT 15

ID 022863 PRELIMINARY; PRT; 756 AA.
 AC 022863;
 DT 01-JAN-1998 (TREMblrel. 05, Created)
 DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)
 DT 01-NOV-1998 (TREMblrel. 08, Last annotation update)
 DE HYPOTHETICAL PROTEIN.
 GN T01024.23.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
 OC Arabidopsis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COLUMBIA;
 RA ROUSLEY S.D., TSCHDY M.M., LIN X., KETCHUM K.A., CROSBY M.L.,
 RA BRANDON R.C., SPRIGGS T.A., MASON T.M., KERLAVAGE A.R., ADAMS M.D.,
 RA SOMERVILLE C.R., VENTER J.C.;
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC002335; AAB64317.1; -
 DR PFAM; PF00566; TBC; 1
 SO SEQUENCE 756 AA; 85182 MW; F71F6CF4 CRC32;

Query Match 10.2%; Score 63.5; DB 10; Length 756;
 Best Local Similarity 34.1%; Pred. No. 66;
 Matches 14; Conservative 7; Mismatches 17; Indels 3; Gaps 1;

QY 21 WGVAVASAVDVIRAAHGEVTEPEARGRLD--ALREWI 58
 DB 355 WLMTLRIYVDVVRTDSHLEFYEDPGNLGRMSDILAVYAVV 395

Search completed: June 23, 2000, 10:07:48
 Job time: 951 sec

Mon Jun 26 09:10:50 2000

us-09-214-478-3.rml

Page 1

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 23, 2000, 09:51:07 ; Search time 42.97 Seconds
(without alignments)
1043.631 Million cell updates/sec

Title: US-09-214-478-3

Perfect score: 1 atgtctctccagctctcc.....cggcgactccgtagag 345

Scoring table: GAPOP 10.0, Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 460926

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: Issued Patents.NA.*

1: /cgn2_6/prodata/1/lna/5/COMB.seq.*
2: /cgn2_6/prodata/1/lna/5/COMB.seq.*
3: /cgn2_6/prodata/1/lna/5/COMB.seq.*
4: /cgn2_6/prodata/1/lna/5/COMB.seq.*
5: /cgn2_6/prodata/1/lna/6/COMB.seq.*
6: /cgn2_6/prodata/1/lna/6/COMB.seq.*
7: /cgn2_6/prodata/1/lna/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query length	ID	Description
1	345	100.0	US-08-735-609-4	Sequence 4, Appl 1
2	345	100.0	US-08-374-483-6	Sequence 6, Appl 1
3	345	100.0	US-08-735-609-1	Sequence 1, Appl 1
4	345	100.0	US-08-379-452-43	Sequence 43, Appl 1
5	80	23.2	US-08-462-014-2	Sequence 2, Appl 1
6	54	15.7	US-08-752-760A-1	Sequence 1, Appl 1
7	29	8.4	US-08-452-267-2	Sequence 2, Appl 1
8	29	8.4	US-08-723-624-18	Sequence 18, Appl 1
9	29	8.4	US-08-630-820-5	Sequence 3, Appl 1
10	29	8.4	US-08-723-624-19	Sequence 19, Appl 1
11	29	8.4	US-08-452-267-3	Sequence 3, Appl 1
12	29	8.4	US-08-318-72A-4	Sequence 4, Appl 1
13	29	8.4	US-08-882-704A-4	Sequence 4, Appl 1
14	29	8.4	US-08-819-866-1	Sequence 1, Appl 1
15	29	8.4	US-09-023-715-1	Sequence 1, Appl 1
16	29	8.4	US-08-819-866-2	Sequence 2, Appl 1
17	29	8.4	US-09-023-715-2	Sequence 2, Appl 1
18	29	8.4	US-08-453-924-4	Sequence 4, Appl 1
19	29	8.0	US-08-471-791-28	Sequence 28, Appl 1
20	29	8.0	PCT-US91-01146-28	Sequence 28, Appl 1
21	27.4	7.9	5432081-1	Patent No. 5432081
22	27.4	7.9	US-08-525-507-14	Sequence 14, Appl 1
23	27.4	7.9	US-08-304-309-3	Sequence 3, Appl 1
24	27.4	7.9	US-08-991-942-3	Sequence 9, Appl 1
25	27.2	7.9	US-08-372-652-9	Sequence 9, Appl 1
26	27.2	7.9	PCT-US95-16311-9	Sequence 9, Appl 1
27	26.8	7.8	US-08-605-106-6	Sequence 6, Appl 1

ALIGNMENTS

28	26.6	7.7	2740	1	US-08-100-247-4	Sequence 4, Appl 1
29	26.6	7.7	2740	2	US-08-483-146A-4	Sequence 4, Appl 1
30	26.6	7.7	2740	2	US-08-232-513A-5	Sequence 5, Appl 1
31	26.6	7.7	2740	2	US-08-484-594A-4	Sequence 4, Appl 1
32	26.4	7.7	2380	7	5268463-1	Patent No. 5268463
33	26.4	7.7	7218	1	US-08-232-463-14	Sequence 14, Appl 1
34	26.2	7.6	34	3	US-08-572-128-2	Sequence 2, Appl 1
35	26.2	7.6	34	4	US-08-757-023A-2	Sequence 2, Appl 1
36	26.2	7.6	1307	4	US-08-960-022-17	Sequence 17, Appl 1
37	26.2	7.6	1849	4	US-08-676-166A-1	Sequence 1, Appl 1
38	26	7.5	852	4	US-08-751-233A-5	Sequence 5, Appl 1
39	25.8	7.5	1320	5	US-08-718-904-76	Sequence 76, Appl 1
40	25.8	7.5	3181	2	US-08-653-086-1	Sequence 1, Appl 1
41	25.6	7.4	295	5	US-08-765-332-167	Sequence 167, Appl 1
42	25.6	7.4	1897	1	US-08-245-688-3	Sequence 3, Appl 1
43	25.6	7.4	1897	1	US-08-245-688-5	Sequence 5, Appl 1
44	25.6	7.4	1897	1	US-08-245-688-7	Sequence 7, Appl 1
45	25.6	7.4	1897	1	US-08-245-688-7	Sequence 7, Appl 1

RESULT 1
US-08-735-609-4/c

Sequence 4, Application US/08735609
Patent No. 595360

GENERAL INFORMATION:

APPLICANT: Chamberlain, Jeffrey S.

APPLICANT: Amalfitano, Andrea

APPLICANT: Hauser, Michael A.

APPLICANT: Kumar-Singh, Rajendra

APPLICANT: Hartigan-O'Connor, Dennis J.

TITLE OF INVENTION: IMPROVED ADENOVIRUS VECTORS

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

ADDRESSER: Medlen & Carroll, LLP

STREET: 220 Montgomery Street, Suite 2200

CITY: San Francisco

STATE: California

COUNTRY: United States of America

ZIP: 94104

COMPUTER READABLE FORM: FLOPPY disk

MEDIUM TYPE: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/735,609

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Ingolia, Diane E.

REGISTRATION NUMBER: 40,027

REFERENCE/DOCKET NUMBER: UM-02484

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 705-8410

TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 34303 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid

DESCRIPTION: /desc = "DNA"

US-08-735-609-4

Query Match 100.0%; Score 345; DB 4; Length 34303;
Best Local Similarity 100.0%; Pred. No. 2.3e-112;
Matches 345; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY	1	agagttcttccagcagctctccgcctccctcccggtgtgagcttcgacagacgaatgtaggt	60
Db	32711	ANGGTTCTCCAGAGCTCTTCCGCTCTCTCCCGTGTGAGATCCGACAGCAATGTGAGGT	32655
OY	61	tggcgtgggtgtgtgcattctctcgagtgatgtattcaaggcagcggcgcataaga	120
Db	32651	TGGCTGGGCTGTGGCTATTCTCCGGTGGTGAATGTTATTAGGGCAGCGCGCATGAAGA	32592
OY	121	gtttacatagaaccccgaaacccgaaggcgccctgtagtctttgagagagtgatactac	180
Db	32591	GTTTTCATATGAACCCGGAAGCCAGGGGGCGCCGTGATGCTTTGAGAGAGTGATATCTAC	32532
OY	181	aactctctacacagagacgatactaaaggcgagaccggagacgcgaagtctttgtcaagc	240
Db	32531	AACCTACTACACAGAGCGATCTTAAGGGGCGAGACCGGAGACGGAGATCTGTTGTACGGC	32472
OY	241	cgacccggtgtttgtcttcaagaaataatgactagctcgccgcttccatttggcatgacct	300
Db	32471	CGCACCTGGTTCCTTACGAGAAATATGACTACATCGTCGGCGCTTCCATTGGCATGCACT	32412
OY	301	acgaccacacagatctggtttcttcgagcgagctccgtaagtag 345	
Db	32411	ACGACCAACAGATCTCGTTCCTTCGCGCACCTCCGTAACAATAAG 32367	

RESULT 2
 US-08-374-483-6/c
 Sequence 6, Application US/08374483
 Patent No. 5880102
 GENERAL INFORMATION:
 APPLICANT: GEORGE, SAMUEL E.
 APPLICANT: BLAZING, MICHAEL A.
 TITLE OF INVENTION: ADENO VIRAL VECTOR SYSTEM
 NUMBER OF SEQUENCES: 9
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: NIXON & VANDERHYTE P. C.
 STREET: 1100 NORTH GLEBE ROAD, 8TH FLOOR
 CITY: ARLINGTON
 STATE: VIRGINIA
 COUNTRY: U.S.A.
 ZIP: 22201-4714
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/374,483
 FILING DATE: 17-JAN-1995
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: WILSON, MARY J.
 REGISTRATION NUMBER: 32,955
 REFERENCE/DOCKET NUMBER: 1579-83
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703) 816-4000
 TELEFAX: (703) 816-4100
 TELEX: 200797 NIXN UR
 INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 34382 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 US-08-374-483-6

```

Query Match      100.0%; Score 345; DB 3; Length 34382;
Best Local Similarity 100.0%; Pred. No. 2.3e-112;
Matches 345; Conservative 0; Mismatches 0; Indels 0;
Gaps 0;
0Y* 1 atggtcttcacagctcttcacgcgtctctcccgctgtgtgactgcgacgaacatgtttagt 60

```

Db	32543	ATGGTTCTTCCACAGCTCTTCCCGCTCCTCCCGGTGTGACTCGCAGAACGATGTAAGT	32484
Oy	61	tgggtgggtgttggtctattctcggtgtgtgtgagtgttatcaagggaagcgcgcatgaaga	120
Db	32483	TGGTGTGGTGTGGCTATTCTTCGGGGTGGATGTATACAGGGCAGGGGGCGATGAAGA	32424
Oy	121	gtttacataagaaccgaagccaaagcgagggcgccgtgattgtctttgagaagatgtacatactaa	180
Db	32423	GTTTACATTAACACCGAACCAGGGGGGCCCTCGATCTGTTGAGAAGTGGATTAATCTAC	32364
Oy	181	aactactaacagagcgatctctaagcggcgagaaacggagaacgagatctgttttgcagcc	240
Db	32363	AACACTACACACAGAGGATCTAAGCGGGAGACCGAGAACGACATCTGTTTGTACGCC	32304
Oy	241	cgacccctggtttgctctcgaagaatatgactcaagtccggcggttccattgtgctgcact	300
Db	32303	CGCACCTGGTTTGGCTTCAGGAATATACATACGCGGGCTTCATTGGCATGACACT	32244
Oy	301	acgacacaacagatcctcggttctcctggcgacatccgttaacatg	345
Db	32243	ACGAGCCAAACGATCTCGGTTTCTTCGGCGCATCTCCGACATG	32199

```

RESULT 3
US-08-735-609-1/c
: Sequence 1, Application US/08735609
: Patent No. 5955360
:
GENERAL INFORMATION:
APPLICANT: Chamberlain, Jeffrey S.
APPLICANT: Amalfitano, Andrea
APPLICANT: Hauser, Michael A.
APPLICANT: Kumar-Singh, Rajendra
APPLICANT: Hartigan-O'Connor, Dennis J.
TITLE OF INVENTION: IMPROVED ADENOVIRUS VECTORS
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Medlen & Carroll, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94104
:
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/735,609
FILING DATE:
:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Ingolia, Diane E.
REGISTRATION NUMBER: 40, 027
REFERENCE/DOCKET NUMBER: UM-02484
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 35935 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
:
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA"
:
US-08-735-609-1

```

Query Match	100.0%;	Score 345;	DB 4;	Length 35935;
Best Local Similarity	100.0%;	Pred. No. 2.3e-112;		
Matches 345;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY 1 atggtctccagctctccgcctccgctggtgacacgagacgaatgtagat 60
|||||
Db 34342 ATGGTCTTCCAGCTCTTCCGCTCCTCCGTGTGACACGAGAACGAATGTAGT 34283
QY 61 tggctggtgtggtctatctcgtggtgtgtatccagggcagcgagcgaatgaaga 120
|||||
Db 34282 TGGCTGGGTGGCTTATCTCGGGTGTGATGTTATCAGGGCAGCGCGCATGAAGA 34223
QY 121 gttacatagaacccagagcagggggcgccctggtatccttgagagagtgatatactac 180
|||||
Db 34222 GTTTACATAGAACCCGAGAGCCAGGGGCGCTGGATGCTTTGAGAGAGTGATATACTAC 34163
QY 181 aactatacagagagagatctaagcggcggagacggagacgagatctgttgcagcc 240
|||||
Db 34162 AACTACTACAGAGCATCTTAAGCGCGAGACCGAGACGAGCATGTTGTGACGCC 34103
QY 241 cgcacctgttctgtctcaggaataatgactacgctcggcgcttccatttgcatgacact 300
|||||
Db 34102 CGCAGCTGGTTTCTTCAGAGAAATATGACTACGTCGGCGTTCATTGGCATGACACT 34043
QY 301 acgacacacagatctcggttgtctcgcgacactcgtaagtag 345
|||||
Db 34042 ACGACCAACAGATCTGGTTGTCTCGCGCACTCCGTACAGTAG 33998

RESULT 4
US-08-379-452-43/c
Sequence 43, Application US/08379452

Patent No. 6040174
GENERAL INFORMATION:
APPLICANT: IMLER, Jean-Luc
APPLICANT: MEHTALI, Majid
APPLICANT: PAVIRANI, Andrea
TITLE OF INVENTION: DEFECTIVE ADENOVIRUSES AND CORRESPONDING
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
STREET: 1737 King Street, Suite 500
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22314-2756
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/379,452
FILING DATE: 26-JAN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/FR94/00624
FILING DATE: 27-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 93 06482
FILING DATE: 28-MAY-1993
ATTORNEY/AGENT INFORMATION:
NAME: Dadio, Susan M.
REGISTRATION NUMBER: 40,373
REFERENCE/DOCKET NUMBER: 029395-002
INFORMATION FOR SEQ ID NO: 43:
SEQUENCE CHARACTERISTICS:
LENGTH: 35935 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-379-452-43

Query Match 100.0%; Score 345; DB 5; Length 35935;
Best Local Similarity 100.0%; Pred. No. 2, 3e-112;
Matches 345; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atggtctccagctctccgcctccgctggtgacacgagacgaatgtagat 60
|||||
Db 34342 ATGGTCTTCCAGCTCTTCCGCTCCTCCGTGTGACACGAGAACGAATGTAGT 34283
QY 61 tggctggtgtggtctatctcgtggtgtgtatccagggcagcgagcgaatgaaga 120
|||||
Db 34282 TGGCTGGGTGGCTTATCTCGGGTGTGATGTTATCAGGGCAGCGCGCATGAAGA 34223
QY 121 gttacatagaacccagagcagggggcgccctggtatccttgagagagtgatatactac 180
|||||
Db 34222 GTTTACATAGAACCCGAGAGCCAGGGGCGCTGGATGCTTTGAGAGAGTGATATACTAC 34163
QY 181 aactatacagagagagatctaagcggcggagacggagacgagatctgttgcagcc 240
|||||
Db 34162 AACTACTACAGAGCATCTTAAGCGCGAGACCGAGACGAGCATGTTGTGACGCC 34103
QY 241 cgcacctgttctgtctcaggaataatgactacgctcggcgcttccatttgcatgacact 300
|||||
Db 34102 CGCAGCTGGTTTCTTCAGAGAAATATGACTACGTCGGCGTTCATTGGCATGACACT 34043
QY 301 acgacacacagatctcggttgtctcgcgacactcgtaagtag 345
|||||
Db 34042 ACGACCAACAGATCTGGTTGTCTCGCGCACTCCGTACAGTAG 33998

RESULT 5
US-08-462-014-2
Sequence 2, Application US/08462014

Patent No. 5756283
GENERAL INFORMATION:
APPLICANT: Wilson, James M.
APPLICANT: Fisher, Krishna J.
APPLICANT: Geo, Guang-ping
TITLE OF INVENTION: Method for Improved Production of
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howson and Howson
STREET: Spring House Corporate Cntr., P.O. Box 457
CITY: Spring House
STATE: PA
COUNTRY: USA
ZIP: 19477
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/462,014
FILING DATE: 05-JUN-1995
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: UPNH1337USA
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9200
TELEFAX: 215-540-5818
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 8299 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: CDNA
US-08-462-014-2

Query Match 23.2%; Score 80; DB 2; Length 8299;
Best Local Similarity 100.0%; Pred. No. 1.8e-18;
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 266 atgactatgctccgagctcatttgagcatcacactacagacacagatctcggttct 335
|||||
DB 3402 ATGACTATGCTCCGAGCTCATTGGCATGACACTACACACAGATCTCGGTCT 3461
|||||
QY 326 cggcgacactccgtacagtag 345
|||||
DB 3462 CGGCGACCTCCGTACAGTAG 3481

RESULT 6
US-08-752-760A-1/C
Sequence 1, Application US/08752760A
Patent No. 5877011

GENERAL INFORMATION:
APPLICANT: Armentano, Donna
APPLICANT: Gregory, Richard J.
APPLICANT: Smith, Alan E.
TITLE OF INVENTION: CHIMERIC ADENOVIRAL VECTORS
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSER: Baker & Botts, L.L.P.
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: NY
COUNTRY: U.S.A.
ZIP: 10112

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/752,760A
FILING DATE: 20-NOV-1996
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER:

ATTORNEY/AGENT INFORMATION:
NAME: Seide, Rochelle K
REGISTRATION NUMBER: 32,300
REFERENCE/DOCKET NUMBER: A31385
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-705-5000
TELEFAX: 212-705-5020
TELEX:

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 35081 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-752-760A-1

Query Match 15.7%; Score 54; DB 3; Length 35081;
Best Local Similarity 52.2%; Pred. No. 7e-09;
Matches 145; Conservative 0; Mismatches 130; Indels 3; Gaps 1;

QY 1 atggtcttcacgagctcccgctctcccggtgtgagctcgacagaaatgtgtaggt 60
|||||
DB 33460 ATGGTCTTCACATCCATCCACCGCCCTCGATGATAGAC--AAGCAGCATTTAC 33404
|||||
QY 61 tggcgtggtgtgagcttctcgtgtgtgtatcagggcgcgcatgaaga 120
|||||
DB 33403 TGGAGGGGAGTGGCTTAAGAGTCTGCTGATGATGAGGGGGAATTCCGATGGACGG 33344
|||||
QY 121 gttacatagaccgacgagcgagggcgctgtgagctgttgaagagagtgatatactac 180
|||||

DB 33343 TTTTATTTCATCAGATGCAGAGAGACTTCCAGAACTTCGAGATGATGACTTTC 33284
QY 181 aactactacagagagcatctaaagcgagacggagacagacagatctgttgcagcc 240
|||||
DB 33283 AGTTGATGACTGAGCGGAGCAGAGGAAAGACGCCCGCAGAGGGGATATATGCTGTCA 33224
QY 241 cgcacgtgttgcgtcaggaatatgactgcgcg 278
|||||
DB 33223 AGGCGCATTTCTGCTGGCAGAGATGACAGAGTAGC 33186

RESULT 7
US-08-452-267-2
Sequence 2, Application US/08452267
Patent No. 5801027

GENERAL INFORMATION:
APPLICANT: Bennett, Malcolm
APPLICANT: May, Sean
APPLICANT: Ramsay, Nicholas
TITLE OF INVENTION: Control of Genes in Transgenic Plants
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSER: Barnes & Thornburg
STREET: 11 South Meridian
CITY: Indianapolis
STATE: IN
COUNTRY: USA
ZIP: 46204

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/452,267
FILING DATE: 26-MAY-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Breen, John P.
REGISTRATION NUMBER: 38,833
REFERENCE/DOCKET NUMBER: 6653-25744
TELECOMMUNICATION INFORMATION:
TELEPHONE: (317) 231-7745
TELEFAX: (317) 231-7433

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2633 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
IMMEDIATE SOURCE:
CLONE: pMIGIT(insert)
US-08-452-267-2

Query Match 8.4%; Score 29; DB 2; Length 2633;
Best Local Similarity 50.4%; Pred. No. 1.3;
Matches 71; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 23 ctctcccggtgtgagctcagacagaaatgtgtaggtgtgaggtgtgtgtattctg 82
|||||
DB 1071 CGCATGTCGGGCAAGATGATACACCGCTGTGTACTGTCAGGTCGTCATGTCG 1130
|||||
QY 83 cgggtgtgtatgtatcagggcgagcgcatgaagagtttcatagaccggaagcca 142
|||||
DB 1131 ATGTCACGCTTGAACATCCGATGATGATGATGATGATGATGATGATGATGAT 1190
|||||
QY 143 gggggcgccgtgagatgcttga 163
|||||
DB 1191 GCGGGACTTTCAGATGTCGA 1211

RESULT 8
US-08-723-624-18

Sequence 18, Application US/08723624
Patent No. 5861277

GENERAL INFORMATION:

APPLICANT: Rose, Alan B.
APPLICANT: Last, Robert L.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR ENHANCING
TITLE OF INVENTION: THE EXPRESSION OF GENES IN PLANTS
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: ARNOLD, WHITE & DUNKEE
STREET: P.O. Box 4433
CITY: Houston
STATE: TX
COUNTRY: USA
ZIP: 77210

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/723,624
FILING DATE: Concurrently Herewith

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Parker, David L.
REGISTRATION NUMBER: 32,165

REFERENCE/DOCKET NUMBER: BRP:002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 512/418-3000
TELEFAX: 512/474-7577

INFORMATION FOR SEQ ID NO: 18:

SEQUENCE CHARACTERISTICS:
LENGTH: 3035 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-723-624-18

Query Match 8.4%; Score 29; DB 3; Length 3035;
Best Local Similarity 50.4%; Pred. No. 1.4; Indels 0; Gaps 0;
Matches 71; Conservative 0; Mismatches 170; Indels 0; Gaps 0;

QY 23 ctctcccggtgtgtactcgcagacgaatgtgtgctggctggctgttcttctg 82

DB 1795 CGCATGCGCGGAAGACTGTAACACGCGCTGTGCTGCGCAGTGGCCCAATGCTG 1854

QY 83 cgggtgtgagttgtatcagggcggcgcatgaagagtttcatagaccggaagcca 142

DB 1855 ATGTACGCGTTGACATGCGGTGATGCGGATCAACAGGTGGTTCAGCAAGCACTA 1914

QY 143 gggggcgctggatgcttga 163

DB 1915 GCGGGACTTTCAGAGTGTGA 1935

RESULT 9

US-08-630-820-5
Sequence 5, Application US/08630820
Patent No. 6008023

GENERAL INFORMATION:

APPLICANT: Oppen, Martin
APPLICANT: Bosslet, Klaus
APPLICANT: Czech, Joerg
TITLE OF INVENTION: CYTOPLASMIC EXPRESSION OF ANTIBODIES,
TITLE OF INVENTION: ANTIBODY FRAGMENTS AND ANTIBODY FRAGMENT FUSION MOLECULES
TITLE OF INVENTION: IN E. COLI
NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESS:

ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/630,820
FILING DATE: 10-APR-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DE 19513676.4
FILING DATE: 11-APR-1995

ATTORNEY/AGENT INFORMATION:

NAME: GRANADOS, Patricia D.
REGISTRATION NUMBER: 33,683

REFERENCE/DOCKET NUMBER: 18748/306
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5390
TELEFAX: (202)672-5399

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 3169 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: circular

MOLECULE TYPE: CDNA

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORGANISM: Enterobacteriaceae: Escherichia coli

STRAIN: PRJ210

IMMEDIATE SOURCE:

CLONE: pTRC99 dicistr. Fab/E.c.-Beta-Gluc

FEATURE:

NAME/KEY: CDS

LOCATION: 3..641

NAME/KEY: CDS

LOCATION: 666..3162

US-08-630-820-5

Query Match 8.4%; Score 29; DB 5; Length 3169;
Best Local Similarity 50.4%; Pred. No. 1.5; Indels 70; Gaps 0;
Matches 71; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 23 ctctcccggtgtgtactcgcagacgaatgtgtgctggctggctgttcttctg 82

DB 1924 CGCATGCGCGGAAGACTGTAACACGCGCTGTGCTGCGCAGTGGCCCAATGCTG 1983

QY 83 cgggtgtgagttgtatcagggcggcgcatgaagagtttcatagaccggaagcca 142

DB 1984 ATGTACGCGTTGACATGCGGTGATGCGGATCAACAGGTGGTTCAGCAAGCACTA 2043

QY 143 gggggcgctggatgcttga 163

DB 2044 GCGGGACTTTCAGAGTGTGA 2064

RESULT 10

US-08-723-624-19
Sequence 19, Application US/08723624
Patent No. 5861277

GENERAL INFORMATION:

APPLICANT: Rose, Alan B.
APPLICANT: Last, Robert L.

Mon Jun 26 09:10:50 2000

us-09-214-478-3.rni

Page 9

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2000, 10:03:34 ; Search time 20.4 Seconds
(Without alignments)

132.364 Million cell updates/sec

Title: US-09-214-478-4

Sequence: 1 MVLPAAPVCDSDQNECVG.....SIWHDITNTISVSAHSVQ 114

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 188963 seqs, 23686106 residues

Total number of hits satisfying chosen parameters: 188963

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: A_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	68	11.0	523 1	P80674 NTRA gene product.
2	63	10.1	394 1	Human ceramide glu
3	63	10.1	394 1	Human HCV repress
4	62.5	10.1	2594 1	IgG-Fc binding pro
5	62.5	10.1	5405 1	IgG-Fc binding pro
6	62	10.0	150 1	Rat myelin oligode
7	61.5	9.9	1661 1	Hydra head activat
8	61.5	9.9	2763 1	Maize dwarf mosaic
9	60	9.7	118 1	Antibody produced
10	60	9.7	118 1	Humanised HmFgl he
11	60	9.7	133 1	Human secreted pro
12	60	9.7	323 1	Human sapiens PSP1
13	60	9.7	377 1	Human sapiens PSP1
14	60	9.7	413 1	Human sapiens PSP1
15	60	9.7	436 1	Human sapiens PSP1
16	60	9.7	458 1	Human sapiens PSP1
17	60	9.7	458 1	Human sapiens PSP1
18	60	9.7	458 1	Human sapiens PSP1
19	60	9.7	458 1	Human sapiens PSP1
20	60	9.7	461 1	Human sapiens PSP1
21	60	9.7	529 1	Human cancer-relat
22	59.5	9.6	343 1	Protein encoded by
23	59.5	9.6	343 1	Mouse prolamine 1
24	59	9.5	176 1	Mouse prolamine 1
25	59	9.5	423 1	Human sapiens PSP1
26	59	9.5	471 1	Human sapiens PSP1
27	59	9.5	649 1	Human sapiens PSP1
28	58.5	9.4	117 1	Human sapiens PSP1
29	58.5	9.4	117 1	Human sapiens PSP1
30	58	9.3	215 1	Human sapiens PSP1
31	57.5	9.3	177 1	Human sapiens PSP1
32	57.5	9.3	571 1	Human sapiens PSP1
33	57	9.2	135 1	Human sapiens PSP1
34	57	9.2	215 1	Human sapiens PSP1

35	57	9.2	264 1	W35562	Human secreted pro
36	57	9.2	274 1	W09819	Human secreted pro
37	57	9.2	521 1	W74802	Human secreted pro
38	57	9.2	532 1	W35565	Human secreted pro
39	57	9.2	1019 1	W43394	Human secreted pro
40	57	9.2	1019 1	W94302	Human secreted pro
41	57	9.2	15281 1	R44929	Human secreted pro
42	56.5	9.1	115 1	R54813	Human secreted pro
43	56.5	9.1	117 1	R53929	Human secreted pro
44	56.5	9.1	3054 1	R40841	Human secreted pro
45	56	9.0	120 1	R13657	Human secreted pro

ALIGNMENTS

RESULT 1	
ID P80674	Standard; protein; 523 AA.
AC P80674	
DT 24-OCT-1990 (first entry)	
DE NTRA gene product.	
KW NTRA, sigma factor; transcriptional activation; nitrogen assimilation;	
KW fixation; nit operon; C4-dicarboxylate transport.	
OS Rhizobium meliloti.	
PN EP-292984-A.	
PD 30-NOV-1988.	
PE 27-MAY-1988; 108482.	
PR 29-MAY-1987; US-055228.	
PA (GENO-) Gen Hospital Corp.	
PI Ronson C, Ausubel F.	
DR WPI: 88-339561/48.	
DR N-PSDB: N81266.	
PT Cloned rhizobium meliloti ntra gene - for use in C4-dicarboxylate	
PT transport, nitrate assimilation, symbiotic nitrogen fixation and	
PT identifying pathogenic genes.	
PS Claim 1; Page 12; 19pp; English.	
CC The ntra gene product is a transcriptional activator. It controls the	
CC processes of nitrate assimilation, symbiotic nitrogen fixation and	
CC C4-dicarboxylate transport, working in conjunction with NtrC, NtrA and	
CC DCD resp. It can also be used to identify genes involved in	
CC pathogenesis.	
CC See also P82126.	
CC Sequence 523 AA;	
QY	Query Match 11.0%; Score 68; DB 1; Length 523;
DB	Best local similarity 28.4%; Pred. No. 5.2; Matches 27; Conservative 20; Mismatches 36; Indels 12; Gaps 4;
QY	23 GVAASAVVAVIRAAAHG--VTEPEARGRLDALREWIYYNTERSKRRDRRSVCH 79
DB	291 GVEAETIDVYVRAAPDGGWLVLPDLPV-----LVNHDYFEIS--RSSRNSGEO 343
QY	80 APTWFCFRKYDVRSIWDITNTISVSAHSVQ 114
DB	344 AFLNECLONAWLTRSL--DQFARTMKVASEIVR 376
RESULT 2	
ID W09088	Standard; Protein; 394 AA.
AC W09088	
DT 03-JUL-1997 (first entry)	
DE Human ceramide glucosyltransferase.	
KW Ceramide glucosyltransferase; GlcT-1; glucosylphingolipid biosynthesis;	
KW glycosylation; glucosylphingoceramide; glucosylceramide; human.	
OS Homo sapiens.	
PN EP-748868-A2.	
PD 18-DEC-1996.	
PF 13-JUN-1996; 109447.	
PR 15-JUN-1995; JP-148472.	
PA (RIKA) INST PHYSICAL & CHEM RES.	

PI Hirabayashi Y, Ichikawa S;
 DR WPI: 97-036152/04.
 DR N-PSDB: T45357.
 PT New isolated human ceramide glucosyl:transferase - useful for prodn.
 PS Claim 1; Page 13; 18pp; English.
 CC The present sequence represents a novel ceramide glucosyltransferase
 CC (gluc-1), which has catalytic activity for glucose transfer from
 CC UDP-glucose to ceramide. Ceramide glucosyltransferase catalyzes the
 CC first glycosylation step of glycosphingoceramide to produce
 CC glucosylceramide, a precursor to more than 300 glycosphingolipid
 CC biosyntheses. The sequence can be used in biochemistry and
 CC biotechnology areas as laboratory agents, and also as preparatory
 CC agents for the manufacture of pharmacological active substances. It
 CC is expected to have use in therapeutic and preventive treatment of
 CC diseases based on reduced expression of ceramide
 CC glucosyltransferase.
 SQ Sequence 394 AA.

Query Match 10.1%; Score 63; DB 1; Length 394;
 Best Local Similarity 19.2%; Pred. No. 14;

Matches 25; Conservative 10; Mismatches 19; Indels 76; Gaps 7;

QY 6 LPAPVCDSDQNEC-----VGLGVAISAVVDYIRAAHEGYIEPEANGRLDAREW 59
 DB 284 LPATITCEPISCEFAVSLIIGW-----AAHH-----VFRMDI 315
 QY 60 YNYTERSKRRDRRRSVCHARTWF-----CPRKYD-----VRRS----- 96
 DB 316 MVEF-----KCHCLAMFLFDYIQLRGVGGTLCESKLDYAVAMFIRESMIT 362
 QY 96 -----TWMDT 100
 DB 363 IFLSALMDPT 372

RESULT 3

ID Y05381 standard; Protein; 394 AA.

AC Y05381;
 DT 30-JUN-1999 (first entry)
 DE Human HCMV repressible gene protein, SEQ ID NO 35.
 KW HCMV repressible gene; Crg; human; human cytomegalovirus; interferon;
 KW anti-viral therapy; anti-HCMV therapy; detection; diagnosis;
 OS Homo sapiens.
 PN MO9913075-A2.
 PD 18-MAR-1999.
 PF 08-SEP-1998; U18638.
 PR 22-SEP-1997; US-059725.
 PR 08-SEP-1997; US-058180.
 PA (UYPR-) UNIV PRINCETON.
 PI Cong J, Schenk T, Zhu H;
 DR WPI: 99-243729/20.
 DR N-PSDB: X33956.
 PT New isolated human genes
 PS Claim 5; Page 174-176; 184pp; English.
 CC This sequence is encoded by a human gene of the invention, repressed in
 CC the presence of HCMV infection, designated HCMV-repressible genes (crg or
 CC crgs). The invention also relates to genes that are induced to express by
 CC both HCMV and interferon (IFN), designated HCMV-inducible genes (cig or
 CC cigs). The products can be used to obtain agents which can
 CC be used for anti-viral therapy, particularly anti-HCMV therapy. They can
 CC also be used for the development of drugs that would allow for higher
 CC dosage IFN treatments without the concomitant toxicity normally
 CC associated with administering high levels of IFN. The products can also
 CC be used for detection, diagnosis and drug screening.
 SQ Sequence 394 AA;

Query Match 10.1%; Score 63; DB 1; Length 394;
 Best Local Similarity 19.2%; Pred. No. 14;

Matches 25; Conservative 10; Mismatches 19; Indels 76; Gaps 7;

QY 6 LPAPVCDSDQNEC-----VGLGVAISAVVDYIRAAHEGYIEPEANGRLDAREW 59
 DB 284 LPATITCEPISCEFAVSLIIGW-----AAHH-----VFRMDI 315
 QY 60 YNYTERSKRRDRRRSVCHARTWF-----CPRKYD-----VRRS----- 96
 DB 316 MVEF-----KCHCLAMFLFDYIQLRGVGGTLCESKLDYAVAMFIRESMIT 362
 QY 96 -----TWMDT 100
 DB 363 IFLSALMDPT 372

RESULT 4

ID W14748 standard; Protein; 2594 AA.

AC W14748;
 DT 13-MAY-1997 (first entry)
 DE IgG-Fc binding protein encoded by 7.8 kb fragment of pNV11-ST.
 KW Fragment 13; pNV11-ST; IgG-Fc binding protein; immunoglobulin; K17;
 KW human; colonic epithelium; monoclonal antibody; K9; probe.
 OS Homo sapiens.
 PN WO9527057-A1.
 PD 12-OCT-1995.
 PF 03-APR-1995; J00638.
 PR 01-APR-1994; JP-129487.
 PR 24-AUG-1994; JP-222547.
 PR 30-MAR-1995; JP-109927.
 PA (CHUS) CHUGAI SEIYAKU KK.
 PI Harada N, Morikawa M;
 DR WPI: 95-358632/46.
 DR N-PSDB: T63073.
 PT DNA derived from colonic epithelium encoding IgG-Fc binding protein
 PT - used in the mapping and analysis of IgG-Fc binding protein mRNA
 PS Claim 1; Page 71-84; 132pp; Japanese.
 CC This sequence is encoded by fragment 13 which is a NotI/KpnI fragment
 CC from pNV11-ST. This sequence represents a portion of the IgG-Fc binding
 CC protein of human colonic epithelium. This sequence was used in the
 CC isolation of the full length sequence given in W14748. mRNA isolated
 CC from human colonic epithelial tissue was used to prepare a cDNA library.
 CC This was screened using monoclonal antibodies K9 and K17 which bind to
 CC the large and small components of the binding protein. Active clones,
 CC see also T63077-81, were used to derive probes for screening a second
 CC DNA library from human colonic epithelial tissue.
 SQ Sequence 2594 AA;

Query Match 10.1%; Score 62.5; DB 1; Length 2594;
 Best Local Similarity 27.7%; Pred. No. 1.6e+02;
 Matches 23; Conservative 7; Mismatches 26; Indels 27; Gaps 3;

QY 5 ALPAPVCDSDQNEC-----VGLGVAISAVVDYIRAAHEGYIEPEANGRLDAREW 57
 DB 1551 ALSAPQC--QDGAEGGCGCGFLXNGACVPIQCGCGHNGYIEPEQVLLD----- 1604
 QY 58 IYNYTERSKRRDRRRSVCH 80
 DB 1604 -----NCRQOCTCHA 1613

RESULT 5

ID W14749 standard; Protein; 5405 AA.

AC W14749;
 DT 13-MAY-1997 (first entry)
 DE IgG-Fc binding protein.
 KW Fragment 13; pNV11-ST; IgG-Fc binding protein; immunoglobulin; K17;
 KW human; colonic epithelium; monoclonal antibody; K9; probe.
 OS Homo sapiens.
 PN WO9527057-A1.
 PD 12-OCT-1995.

PF 03-APR-1995; J00638.
 PR 01-APR-1994; JP-129487.
 PR 24-AUG-1994; JP-222547.
 PR 30-MAR-1995; JP-109927.
 PA (CHUS) CHUGAI SEIYAKU KK.
 PI Harada N, Morikawa M;
 DR WPI; 95-358632/46.
 DR N-PSDB; T63074.
 PT DNA derived from colonic epithelium encoding IgG-Fc binding protein
 PT - used in the mapping and analysis of IgG-Fc binding protein mRNA
 PS Claim 3; Page 86-113; 132pp; Japanese.
 CC This sequence represents the IgG-Fc binding protein of human colonic
 CC epithelium. This sequence was isolated using the sequence given in
 CC W44748. mRNA isolated from human colonic epithelial tissue was used
 CC to prepare a cDNA library. This was screened using monoclonal antibodies
 CC K9 and K17 which bind to the large and small components of the binding
 CC protein. Active clones, see also T63077-81, were used to derive probes
 CC for screening a second DNA library from human colonic epithelial tissue.
 SQ Sequence 5405 AA;

Query Match 10.1%; Score 62.5; DB 1; Length 5405;
 Best Local Similarity 27.7%; Pred. No. 4e+02; Mismatches 26; Indels 27; Gaps 3;
 Matches 23; Conservative 7;

OY 5 ALPAPVCDSONECV-----GWLGVAVSAVVDVIRAAHEGYIEPEARGLDALREW 57
 DB 1551 ALSAPPC- QDGCAGCGCCSGFLYNGACVPIQCCGCHNGYVEEGRVLID----- 1604
 OY 58 IYNYIYERSKRDRRRRSVCHA 80
 DB 1604 -----NCRQCCTCHA 1613

RESULT 6
 ID W97817
 AC W97817 standard; Protein; 150 AA.
 DT 21-MAY-1999 (first entry)
 DE Rat myelin oligodendrocyte glycoprotein.
 KW Myelin oligodendrocyte glycoprotein; MOG; rat; butyrophilin;
 KW multiple sclerosis; diagnosis; vaccine; transgenic animal; food;
 KW dairy product; milk.
 OS Rattus sp.
 PN W09905162-A1.
 PD 04-FEB-1999
 PR 22-JUL-1998; U15062.
 PR 22-JUL-1997; US-053376.
 PA (UYMA-) UNIV MARYLAND BALTIMORE.
 PI Llanington C, Mather IH;
 DR WPI; 99-142845/12.
 PT New modified butyrophilin - not cross-reactive with myelin
 PT oligo-dendrocyte glycoprotein, useful in dairy products, vaccines,
 PT and assays for susceptibility to multiple sclerosis
 PS Disclosure: Page 46-47; 53pp; English.
 CC This polypeptide comprises the rat myelin oligodendrocyte
 CC glycoprotein (MOG). The invention relates to butyrophilin (BTN,
 CC see W97812-16), a major component of the milk fat globule membrane
 CC and member of the immunoglobulin (Ig) superfamily. The intermediate
 CC -type (Ig1) domain of BTN has a high degree of identity with the IgV
 CC domain of MOG. Immunisation with exogenous BTN induces the
 CC production of antibodies which cross-react with endogenous MOG, and
 CC this predisposes animals to pathological autoimmune responses such
 CC as multiple sclerosis (MS). The invention provides methods of
 CC modifying BTN to eliminate sequences that elicit this immune
 CC response to MOG, e.g. by removing or modifying the Ig1 domain of a
 CC non-human BTN (nbBTN) in a dairy product, or by producing a
 CC humanised chimeric BTN. The invention includes genetic manipulation
 CC of the endogenous BTN gene of an animal such that the human Ig1
 CC domain is introduced into the germline of the animal, and the
 CC animal expresses the humanised protein in its milk. Methods are
 CC provided for detecting susceptibility to MS in a human patient, from
 CC ingesting, and for identifying a human patient at risk of developing

CC a pathogenic autoimmune response to MOG on exposure to nbBTN is also
 CC possible, by screening the patient to identify polymorphisms in the
 CC gene with domains which are at least 50% identical to nbBTN Ig1
 CC domains (claimed).
 SQ Sequence 150 AA;

Query Match 10.0%; Score 62; DB 1; Length 150;
 Best Local Similarity 37.5%; Pred. No. 5.8;
 Matches 15; Conservative 5; Mismatches 20; Indels 0; Gaps 0;

OY 19 VGMGVAVSAVVDVIRAAHEGYIEPEARGLDALREW 58
 DB 37 VGMVRSFSPSRVHLRNGKDDAQAPEYGRTELKESI 76

RESULT 7
 ID W76041
 AC W76041 standard; Protein; 1661 AA.
 DT 20-NOV-1998 (first entry)
 DE Hydra head activator binding protein.
 KW Head activator binding protein; hydra; screening assay; agonist;
 KW antagonist; treatment; neurodegenerative disease.
 OS Chironhydra viridissima.
 PS Key Location/Qualifiers
 FT Peptide 1..29
 FT /label= signal
 FT 30..1661
 FT /label= head_activating_binding_protein
 FT 80..84
 FT /note= "protease site"
 FT 85..99
 FT /label= N-terminus
 FT 121..737
 FT /label= VPS10
 FT /note= "As given in specification"
 FT 738..1002
 FT /label= 6XB
 FT /note= "As given in specification"
 FT 1003..1041
 FT /note= "EGF-like region as given in the specification"
 FT 1042..1367
 FT /label= 7XA
 FT /note= "As given in specification"
 FT 1368..1543
 FT /note= "Fibronectin type III"
 FT 1584..1606
 FT Domain /label= transmembrane
 PN DE19808258-A1.
 PD 03-SEP-1998.
 PD 27-FEB-1998; 008258.
 PR 10-JUL-1997; DE-028490.
 PR 28-FEB-1997; DE-008134.
 PA (EVOT-) EVOTEC BIOSYSTEMS GMBH.
 PI Hampe W, Schaller C;
 DR WPI; 98-468562/41.
 DR N-PSDB; V46335.
 PT Hydra head activator binding protein - useful in screening assay for
 PT agonists or antagonists
 PS Claim 3; Page 6-11; 21pp; German.
 CC This sequence represents a head activator binding protein isolated from
 CC Chironhydra viridissima by chromatographic methods. The protein is used
 CC in a screening assay for agonists or antagonists comprising determining
 CC the binding properties of candidate substances with the protein. The
 CC assay can be used to screen for substances useful for treating
 CC neurodegenerative diseases.
 SQ Sequence 1661 AA;

Query Match 9.9%; Score 61.5; DB 1; Length 1661;
 Best Local Similarity 31.3%; Pred. No. 1.2e+02;
 Matches 21; Conservative 4; Mismatches 25; Indels 17; Gaps 3;

```

OY      8 APPVCDSONECVWTL-----GVAYSAVVDYIRAAHEGYIIEPARGRLDALR-----E 56
DB      499 SPSVYFSSNAGISMKRIPGNTYIAVD-----HGVIVGVKIGMTSYLNRSDGNT 552
OY      57 WIIYNYT 63.
DB      553 WYSYNYT 559

RESULT  8
W10344  W10344 standard; Protein; 2763 AA.
AC      W10344;
DE      15-APR-1997 (first entry)
KW      MDW-B; viral resistance; disease resistance; transgenic plant;
KW      monococt; P3 proteinase; N1a proteinase; N1b replicase;
KW      RNA-dependent RNA polymerase; coat protein; Zea mays; sorghum;
KW      sugarcane; Saccharum officinale.
OS      Maize dwarf mosaic virus strain B.
FH      Key
FH      Location/Qualifiers
FT      1..377
FT      /label= "HC-Pro
FT      /note= "portion of the helper component-P2
FT      proteinase"
FT      263..336
FT      /note= "MDW-B HC-Pro domain"
FT      378..791
FT      /label= "P3-proteinase
FT      /note= "claimed polypeptide (claim 20)"
FT      792..1430
FT      /label= "Cylindrical inclusion protein
FT      /note= "claimed polypeptide (claim 20)"
FT      880..1010
FT      /note= "conserved helicase domain"
FT      1431..1483
FT      /label= "K2
FT      1484..1914
FT      /note= "6 kDa protein"
FT      /label= "N1a-proteinase
FT      /note= "claimed polypeptide (claim 20)"
FT      1915..2435
FT      /label= "N1b-replicase
FT      /note= "claimed polypeptide (claim 20)"
FT      2266..2268
FT      /note= "conserved motif characteristic of
FT      RNA-dependent replicases"
FT      2436..2763
FT      /label= "Coat-protein
FT      144
FT      /note= "unidentified amino acid"
FT      704
FT      /note= "unidentified amino acid"
FT      712
FT      /note= "unidentified amino acid"
FT      829
FT      /note= "unidentified amino acid"
FT      834
FT      /note= "unidentified amino acid"
FT      843
FT      /note= "unidentified amino acid"
FT      847
FT      /note= "unidentified amino acid"
FT      852
FT      /note= "unidentified amino acid"
FT      1346
FT      /note= "unidentified amino acid"
FT      1362
FT      /note= "unidentified amino acid"
FT      2077
FT      /note= "unidentified amino acid"
FT      misc_difference 2385

```

```

FT      /note= "unidentified amino acid"
PN      MO9702352-A1.
PD      23-JAN-1997.
PF      20-JUN-1996; E02673.
PR      30-JUN-1995; US-496944.
PA      (CIBA ) CIBA GEIGY AG.
PI      Dietz JM, Law MD;
DR      WPI: 97-108965/10.
DR      N-PSDB; T47073.
FT      Chimeric gene for imparting viral resistance to plants - contains
FT      sequence modified to express non-translatable mRNA, or non-coat
FT      viral protein
PS      Claim 20; Page 31-44; 64pp; English.
CC      The sequence of the polypeptide encoded by the polyclonistic mRNA
CC      (T47073) of maize dwarf mosaic virus strain B (MDW-B) is given in
CC      W10344. New chimeric genes comprise a monocotyledonous plant
CC      promoter linked to a modified nucleic acid sequence derived from
CC      the MDW-B genome. The modification is such that mRNA is translated
CC      to a truncated protein (pref. smaller than 200 amino acids), no
CC      translation of mRNA occurs or the transcribed mRNA lacks the
CC      translation initiation codon or includes a premature stop codon.
CC      Expression of the chimeric gene inhibits infection of plants (pref.
CC      sorghum, sugarcane, esp. maize) by MDW. The transgenic plants
CC      display an inheritable resistance trait.
SQ      Sequence 2763 AA;

Query Match      9.9%; Score 61.5; DB 1; Length 2763;
Best Local Similarity 29.8%; Pred. No. 2.3e+02;
Matches 14; Conservative 9; Mismatches 17; Indels 7; Gaps 1;

OY      34 RAAHEGYIIEPARGRLDALRENYTYRERKRRRRRSYCHA 80
DB      2313 RGKTEGIIYFKERKIVALEW-----DRSNLPBRLAICAA 2352

RESULT  9
R22421  R22421 standard; Protein; 118 AA.
AC      R22421;
DE      29-JUL-1992 (first entry)
KW      Antibody produced from HuVH1conHMFg1.
KW      Anti-HMFg specificity; complementarity determining region; CDR;
KW      human milk fat globule; polymorphic epithelial mucin; PEM.
OS      Homo sapiens.
FH      Key
FH      Location/Qualifiers
FT      31..35
FT      /label= "CDR1
FT      /note= "from mouse heavy chain"
FT      50..65
FT      /label= "CDR2
FT      /note= "from mouse heavy chain"
FT      95..103
FT      /label= "CDR3
FT      /note= "from mouse heavy chain"
FT      region
FT      MO9204380-A.
FT      19-MAR-1992.
FT      05-SEP-1991; G01511.
FT      07-SEP-1990; GB-019553.
FT      (UNIL ) UNILEVER PLC.
FT      Verhoeven M;
FT      WPI: 92-114305/14.
FT      N-PSDB: 023349.
DR      Synthetic specific binding agent and reshaped human antibody -
FT      specific for human polymorphic epithelial mucin for treatment and
FT      in-vivo diagnosis of PEM-producing cancers
PS      Claim 11; Fig 12; 62pp; English.
CC      The reshaped antibody sequence was deduced from the sequence of a
CC      reshaped human variable heavy chain gene which had a human class I
CC      framework but contained mouse CDRs. To reshape a human heavy chain,
CC      three fragments, each one contg. a murine CDR, were assembled to
CC      form gene HuVH1conHMFg1. The reshaped gene was inserted into a
CC      pSV vector to produce a plasmid for transformation of E. coli.

```

CC The antibodies produced are essentially human so are unlikely
CC to cause adverse reactions. The reshaped molecule can be used
CC to treat PEM-producing cancers (e.g. of the breast, ovary, uterus
CC or lung) and in in-vivo diagnosis.
CC See also R22419-22.
SQ Sequence 118 AA:

Query Match 9.7%; Score 60; DB 1; Length 118;
Best Local Similarity 23.2%; Pred. No. 7.4;
Matches 23; Conservative 15; Mismatches 39; Indels 22; Gaps 4;

QY 23 GVAYSANVDVIRAAHNECV-----YIEPARGRLALREMIYNYTTERS 67
DB 26 GTTFSAYWIEWRQAGKLEWVGEILPGSNNSRYNE-KFKGRVYTRDTSTNTAYMELS 84

QY 68 KRDRRRRSVCHARKWFCRKDYVRSIWHDTTNTIS 106
DB 85 SLRSE-----DTAVYCARSDYDFAMFAWGGTLVTVS 117

RESULT 10

RS4758 ID RS4758; standard; Protein; 118 AA.

AC RS4758 (first entry)

DE Humanised HMFGL heavy chain variable region.

KW Single chain Fv fragment; Scfv; Gene Therapy; Monoclonal Antibody;

OS Human Milk Fat Globule Antigen; HMFGL.

OS Mus sp

OS Homo sapiens.

FT Key Location/Qualifiers

FT misc_difference 5 /label- humanised by replacing Gln with Val

FT misc_difference 11.12 /label- humanised by replacing Leu and Met with

FT misc_difference 20 /label- humanised by replacing Ile with Val

FT misc_difference 25 /label- humanised by replacing Thr with Ser

FT misc_difference 38 /label- humanised by replacing Lys with Arg

FT misc_difference 40 /label- humanised by replacing Arg with Ala

FT misc_difference 43 /label- humanised by replacing His with Lys

FT misc_difference 48 /label- humanised by replacing Ile with Val

FT misc_difference 67.68 /label- humanised by replacing Lys and Ala

FT misc_difference 70 with Arg and Val

FT misc_difference 72 /label- humanised by replacing Phe with Val

FT misc_difference 76 /label- humanised by replacing Ala with Arg

FT misc_difference 82 /label- humanised by replacing Ser with Thr

FT misc_difference 87 /label- humanised by replacing Gln with Glu

FT misc_difference 91 /label- humanised by replacing Thr with Arg

FT misc_difference 97 /label- humanised by replacing Ser with Thr

FT misc_difference 113 /label- humanised by replacing Ser with Ala

FT misc_difference 118 /label- humanised by replacing Pro with Leu

FT protein 1.118 /label- humanised by replacing Ala with Ser

FT region 1.30 /note- "heavy chain variable region"

FT region /label- FR 1
FT region 31.35 /label- CDR 1
FT region 36.49 /label- FR 2
FT region 50.66 /label- CDR 2
FT region 67.98 /label- FR 3
FT region 99.107 /label- CDR 3
FT region 108.118 /label- FR 4

FT region /label- FR 4

MO9410323-A.
PD 11-MAY-1994.
PD 04-NOV-1993: G02267.
PR 04-NOV-1992: GB-023084.
PA (IMCR) IMPERIAL CANCER RES TECHNOLOGY.
PI Epineuros AA, Spooner RA;
DR N-PSDB; 064819.
DR WPI; 94-167477/20.
PT Virus with modified binding moiety specific for the target cells
PT - used to deliver genes for gene therapy and cancer treatment

PS Example 4; Page 74; 110pp; English.
CC RS4758 shows an humanised HMFGL heavy chain variable region. The
CC mouse HMFGL is changed at specific residues within the framework region.
CC It is used to generate an scfv with the humanised HMFGL light chain
CC variable region (R63804).
CC The invention is concerned with the fusion of an scfv with the Ad5
CC fibre such that the fusion sequence can be used to target cells for
CC gene therapy.
SQ Sequence 118 AA:

Query Match 9.7%; Score 60; DB 1; Length 118;
Best Local Similarity 23.2%; Pred. No. 7.4;
Matches 23; Conservative 15; Mismatches 39; Indels 22; Gaps 4;

QY 23 GVAYSANVDVIRAAHNECV-----YIEPARGRLALREMIYNYTTERS 67
DB 26 GTTFSAYWIEWRQAGKLEWVGEILPGSNNSRYNE-KFKGRVYTRDTSTNTAYMELS 84

QY 68 KRDRRRRSVCHARKWFCRKDYVRSIWHDTTNTIS 106
DB 85 SLRSE-----DTAVYCARSDYDFAMFAWGGTLVTVS 117

RESULT 11

Y13128 ID Y13128 standard; Protein; 133 AA.

AC Y13128; 22-JUN-1999 (first entry)

DE Human secreted protein encoded by 5' EST SEQ ID NO: 142.

KW Human; secreted protein; EGF; expressed sequence tag; diagnosis;

KW forensic; gene therapy; chromosome mapping; signal peptide;

KW upstream regulatory sequence; cytokine activity; cell proliferation;

KW differentiation; haematopoiesis regulation; tissue growth regulation;

KW reproductive hormone regulation; chemotactic; haemostatic;

KW thrombolytic; anti-inflammatory; tumour inhibition.

OS Homo sapiens.

PN WO906552-42.

PD 11-FEB-1999.

PR 31-JUL-1998; IB1236.

PR 01-AUG-1997; US-9052223.

PA (GEST) GENSET.

PI Ducleit A, Dumas Mline Edwards J, Lacroix B;

DR WPI; 99-153782/13.

DR N-PSDB; X51928.

PT New isolated brain-derived nucleic acids - used to develop products

PT which may have cytokine, immune, regulatory, haematopoiesis

PS regulating, anti-inflammatory or tumour inhibition activity

PS Claim 34; Page 522-523; 577pp; English.

PS X51767 to X52019 represent 5' expressed sequence tags (ESTs) for human

secreted proteins, and encode the proteins given in Y12987 to Y13219, respectively. The proteins given represent the signal peptide and an N-terminal fragment of a secreted protein. The nucleic acid sequences can be used for producing secreted human gene products. They can also be used to develop products for diagnosis and therapy. The proteins obtained may have cytokine activity, cell proliferation/differentiation activity, haematopoiesis regulating activity, tissue growth regulating activity, reproductive hormone regulating activity, chemotactic/chemokinetic activity, haemostatic and thrombolytic activity, receptor/ligand activity, anti-inflammatory activity, tumour inhibition activity or other activities. The products can be used in forensic, gene therapy and chromosome mapping procedures. The sequences can also be used for obtaining corresponding promoter sequences. The nucleic acids encoding the signal peptide can be used for directing extracellular secretion of a polypeptide or the insertion of a polypeptide into a membrane, or importing a polypeptide into a cell.

Sequence 133 AA;

Query Match 9.7%; Score 60; DB 1; Length 133;
Best Local Similarity 41.7%; Pred. No. 8.6;
Matches 10; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 60 YNYTERSKRRDRRRSVCARTW 83
DB 32 YAYHERKERDPRHACLDGFW 55

RESULT 12

W56774 standard; Protein; 323 AA.

AC W56774; 13-OCT-1998 (first entry)

DE Homo sapiens PSP1 partial sequence.

KW PS-1; presenilin; presenilin-1; PSP-1; Alzheimer's disease;

OS serine protease; neurodegeneration; predisposition; diagnosis.

PN EP-828003-A2.

PD 11-MAR-1998.

PF 26-AUG-1997; 306501.

PR 13-DEC-1996; US-032875.

PR 06-SEP-1996; US-025436.

PR 25-OCT-1996; US-027873.

PA (SMIR) SMITHKLINE BEECHAM CORP.

PI Browne MJ, Clinkenbeard HE, Creasy CL, Karan EH,

PI Livi GP, Southern CD;

DR WPI; 98-161101/15.

DR N-PSDB; V29522.

PT Nucleic acids encoding human serum protease protein(s) - used for

diagnosing pre-disposition to Alzheimer's disease, etc.

PS Example 2; Page 18-19; 65pp; English.

CC The sequence is that of at least a fragment of the serine

protease PSP1. This can be used to identify modulators

of serine protease activity and also to diagnose a

condition associated with lack of one of the serine proteases

or a genetic predisposition to neurodegeneration in a patient,

preferably predisposition to Alzheimer's disease.

Sequence 323 AA;

Query Match 9.7%; Score 60; DB 1; Length 323;
Best Local Similarity 32.6%; Pred. No. 25;
Matches 14; Conservative 8; Mismatches 13; Indels 8; Gaps 1;

QY 2 VLPALPAPVCSQNECVGMLGVAVSAYVDVIRAAHGEVYLE 44
DB 77 VLAAYVSPPPASPRSQ-----YNFADVVEKTAIPAVVYLE 111

RESULT 13
W56774 standard; Protein; 377 AA.

AC W56774;
DT 13-OCT-1998 (first entry)
DE Homo sapiens PSP1-3 consensus sequence.
KW PS-1; presenilin; presenilin-1; PSP-1; Alzheimer's disease;
OS serine protease; neurodegeneration; predisposition; diagnosis.
PN EP-828003-A2.

PD 11-MAR-1998.

PF 26-AUG-1997; 306501.

PR 13-DEC-1996; US-032875.

PR 06-SEP-1996; US-025436.

PR 25-OCT-1996; US-027873.

PA (SMIR) SMITHKLINE BEECHAM CORP.

PI Browne MJ, Clinkenbeard HE, Creasy CL, Karan EH,

PI Livi GP, Southern CD;

DR WPI; 98-161101/15.

DR N-PSDB; V29537.

PT Nucleic acids encoding human serum protease protein(s) - used for

diagnosing pre-disposition to Alzheimer's disease, etc.

PS Claim 7; Page 41-42; 65pp; English.

CC The sequence is that of the consensus sequence of PSP1-3

serine protease which can be used to identify modulators of

serine protease activity and also to diagnose a condition

associated with lack of one of the serine proteases or

a genetic predisposition to neurodegeneration in a patient,

preferably predisposition to Alzheimer's disease.

Sequence 377 AA;

Query Match 9.7%; Score 60; DB 1; Length 377;
Best Local Similarity 32.6%; Pred. No. 31;
Matches 14; Conservative 8; Mismatches 13; Indels 8; Gaps 1;

QY 2 VLPALPAPVCSQNECVGMLGVAVSAYVDVIRAAHGEVYLE 44
DB 131 VLAAYVSPPPASPRSQ-----YNFADVVEKTAIPAVVYLE 165

RESULT 14

ID R13506 standard; Protein; 413 AA.

AC R13506;

DT 25-OCT-1991 (first entry)

DE P.denitrificans COB L.

KW cob gene; corrinoid; descobalocorrinoid; cor gene.

OS Pseudomonas denitrificans.

PN WO911518-A.

PD 08-AUG-1991.

PF 30-JAN-1991; F00054.

PR 31-JAN-1990; FR-001137.

PA (RHON) RHONE-POULENC BIOCH.

PI Blanche F, Meron B, Crouzet J, Debussche L, Levy-Schli S;

PI Thibaut D;

DR WPI; 91-252650/34.

DR N-PSDB; Q13285.

PT New polypeptide(s) involved in cobalamin and cobamide

biosynthesis - and DNA encoding them, for amplification of

PT cobalamin, esp. coenzyme B12 prodn.

PS Claim 17; Fig 16; 289pp; French.

CC This sequence corresponds to one of 24 polypeptides obtained from

P.denitrificans and implicated in the biosynthesis of cobalamines

and/or cobamides, specifically in catalysing the transfer of a

methyl group to positions C1, C5, C11, C15 or C17 in the conversion

of precorrin-3 to cobyrinic acid a,c-diamide. It is encoded by part

of the 8.7kb EcoRI-EcoRI fragment of plasmid pX1367. The plasmid was

isolated from a P.denitrificans genomic DNA bank constructed in

vector pX159.

CC See Q13284-Q13288.

CC Sequence 413 AA;

Query Match 9.7%; Score 60; DB 1; Length 413;

Best Local Similarity 48.6%; Pred. No. 34;
Matches 17; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

QY 22 LGVAYSADVIRAAHEGYIEPEARGLDALRE 56
DB 205 LGIVHPLNVCALIEVADEGARILPLAAGRDDALFE 239

RESULT 15

W56776
ID W56776 strand; Protein; 436 AA.
AC W56776;
DT 13-OCT-1998 (first entry)
DE Homo sapiens PSP1-4 consensus sequence.
KW PS-1; presenilin; presenilin-1; PSP1-4; Alzheimer's disease;
KW serine protease; neurodegeneration; predisposition; diagnosis.
OS Homo sapiens.
PN EP-828003-A2.
PD 11-MAR-1998.
PF 26-AUG-1997; 306501.
PR 13-DEC-1996; US-032875.
PR 06-SEP-1996; US-025436.
PR 25-OCT-1996; US-027873.
PA (SMIR) SMITHKLINE BEECHAM CORP.
PI Browne MJ, Clinkenbeard HE, Creasy CL, Karan EH,
PI Livi GP, Southan CD;
DR WPI; 98-161101/15.
DR N-PSDB; V29538.
PT Nucleic acids encoding human serum protease protein(s) - used for
PS diagnosing pre-disposition to Alzheimer's disease, etc.
PS Claim 7; Page 44-45; 65pp; English.
CC The sequence is that of the consensus sequence of PSP1-4
CC serine protease which can be used to identify modulators of
CC serine protease activity and also to diagnose a condition
CC associated with lack of one of the serine proteases or
CC a genetic predisposition to neurodegeneration in a patient,
CC preferably predisposition to Alzheimer's disease.
SQ Sequence 436 AA;

Query Match 9.7%; Score 60; DB 1; Length 436;
Best Local Similarity 32.6%; Pred. No. 37;
Matches 14; Conservative 8; Mismatches 13; Indels 8; Gaps 1;

QY 2 VLPALPAPVCDSONECVGLGVAYSADVIRAAHEGYIE 44
DB 131 VLAAVPSPPPASPRSQ-----YNFADVVEKTAAPAVYIE 165

Search completed: June 23, 2000, 10:03:38
Job time: 3366 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using SW model

Run on: June 23, 2000, 10:04:33 ; Search time 45.41 Seconds
(without alignments)
36.245 Million cell updates/sec

Title: US-09-214-478-4

Perfect score: 621
Sequence: 1 MYLPALPAPPCDSQNECVG.....SIMHDTTNTISVSASHSYQ 114

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 145308 seqs, 14437401 residues

Total number of hits satisfying chosen parameters: 145308

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

1: Issued Patents, AA:*
2: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	63	10.1	394	1	US-08-663-713A-2
2	61.5	9.9	2763	3	US-08-496-944-2
3	60	9.7	118	2	US-08-428-257A-78
4	59	9.5	649	3	US-08-996-441B-109
5	58.5	9.4	890	2	US-08-483-101-14
6	57	9.2	135	2	US-08-860-174A-4
7	57	9.2	215	2	US-08-659-251-8
8	57	9.2	215	4	PCT-US96-11445-8
9	57	9.2	274	2	US-08-860-174A-12
10	57	9.2	1019	1	US-08-296-014A-4
11	57	9.2	1019	2	US-08-596-405-4
12	57	9.2	1019	2	US-08-877-620-4
13	57	9.2	15281	2	US-08-471-119A-2
14	56.5	9.1	115	2	US-08-379-057-31
15	56.5	9.1	115	2	US-08-428-197-42
16	56.5	9.1	115	4	PCT-US93-10555-42
17	56.5	9.1	788	2	US-08-918-914-4
18	56	9.0	622	2	US-08-458-146-2
19	56	9.0	622	2	US-08-459-065-2
20	56	9.0	1015	1	US-08-537-210A-1
21	56	9.0	2471	1	US-08-185-432-16
22	56	9.0	2471	1	US-08-083-590A-19
23	55.5	8.9	137	1	US-08-331-398A-61
24	55.5	8.9	137	2	US-08-331-387B-61
25	55.5	8.9	137	2	US-08-753-804A-60
26	55	8.9	247	4	PCT-US94-10257A-2
27	55	8.9	489	4	PCT-US96-10521-9
28	55	8.9	503	2	US-08-394-189B-2
29	55	8.9	503	4	PCT-US93-05701-19

30	55	8.9	503	4	PCT-US93-05705-2	Sequence 2, Appl1
31	54.5	8.8	487	1	US-08-249-112-4	Sequence 4, Appl1
32	54.5	8.8	487	1	PCT-US95-06556-4	Sequence 4, Appl1
33	54.5	8.8	611	1	US-08-386-727-4	Sequence 4, Appl1
34	54.5	8.8	611	2	US-08-600-452A-4	Sequence 4, Appl1
35	54	8.7	138	2	US-08-379-057-14	Sequence 14, Appl1
36	54	8.7	144	2	US-08-116-778E-36	Sequence 36, Appl1
37	54	8.7	144	2	US-08-438-562-36	Sequence 36, Appl1
38	54	8.7	144	2	US-08-483-528B-100	Sequence 100, Appl1
39	54	8.7	379	2	US-08-871-074-26	Sequence 26, Appl1
40	54	8.7	489	1	US-08-489-733-6	Sequence 6, Appl1
41	54	8.7	489	2	US-08-993-581B-6	Sequence 6, Appl1
42	54	8.7	491	1	US-08-489-733-5	Sequence 5, Appl1
43	54	8.7	491	2	US-08-993-581B-5	Sequence 5, Appl1
44	54	8.7	760	1	US-08-230-491A-2	Sequence 2, Appl1
45	54	8.7	760	1	US-08-519-280A-2	Sequence 2, Appl1

ALIGNMENTS

RESULT 1
US-08-663-713A-2
; Sequence 2, Application US/08663713A
; Patent No. 5780284
; GENERAL INFORMATION:
; APPLICANT: YOSHIO HIRABAYASHI et al.
; TITLE OF INVENTION: CERAMIDE GLUCOSYLTRANSFERASE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 Inch, 500 kb
; OPERATING SYSTEM: IBM Compatible
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/663,713A
; FILING DATE: June 14, 1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX:
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 394 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-663-713A-2

Query Match 10.1% Score 63; DB 1; Length 394;
Best Local Similarity 19.2%; Pred. No. 6.8; Indels 76; Gaps 7;
Matches 25; Conservative 10; Mismatches 19; Indels 76; Gaps 7;
QY 6 LPAPPCDSQNECVG-----VGMGLVAASAVVDVIRAAAHGCVITIEPARGRLDALREWIY 59
DB 284 LPATPCIPICPCFVASLIQM-----AAHH-----VFRNDI 315

OY 60 YNTERSKRRDRRRSVCARTWF-----CFKRYD-----VRS----- 96
DB 316 MYFE-----MCHCLAMFIDYIQLNGVGGTLCFSKLDYAVAMFRESNTIY 362
OY 96 -----IWHDT 100
DB 363 IFLSALMDPT 372

RESULT 2

US-08-496-944-2
Sequence 2, Application US/08496944
Patent No. 6040496

GENERAL INFORMATION:

APPLICANT: Law, Marcus D
APPLICANT: Dietz, Jon M
TITLE OF INVENTION: Use of Translationally altered RNA to
TITLE OF INVENTION: Confer Resistance to Maitre Dwarf Mosaic Virus and Other
TITLE OF INVENTION: Monocytiledonous Plant Viruses
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESSES:
ADDRESSEE: CIBA-Geigy Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA

ZIP: 10532

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30B

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/496,944

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Elmer, James Scott

REGISTRATION NUMBER: 36,129

REFERENCE/DOCKET NUMBER: CGC 1814

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 2763 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-496-944-2

Query Match 9.7%; Score 61.5; DB 3; Length 2763;
Best Local Similarity 29.8%; Pred. No. 1.3e+02;
Matches 14; Conservative 9; Mismatches 17; Indels 7; Gaps 1;

OY 34 RAAAHGVYIEPARGRLALREWIYNYTERSKRRDRRRSVCCHA 80
DB 2313 KGIKYGIIIPKIKERIVAILDEM-----DSSNLPERRLEAICA 2352

RESULT 3

US-08-428-257A-78
Sequence 78, Application US/08428257A
Patent No. 5885808

GENERAL INFORMATION:

APPLICANT: Spooner, Robert A.
APPLICANT: Epenetos, A.A.

TITLE OF INVENTION: Compounds to target cells
NUMBER OF SEQUENCES: 80

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Jules E. Goldberg

STREET: 261 Madison Avenue

CITY: New York

STATE: NY

COUNTRY: USA

ZIP: 10016-2391
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/428,257A
FILING DATE: 07/05/95
CLASSIFICATION: 514
INFORMATION FOR SEQ ID NO: 78:
SEQUENCE CHARACTERISTICS:
LENGTH: 118 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-428-257A-78

Query Match 9.7%; Score 60; DB 2; Length 118;
Best Local Similarity 23.2%; Pred. No. 3.3;
Matches 23; Conservative 15; Mismatches 39; Indels 22; Gaps 4;

OY 23 GVAISAV-VDVIRAAAHGVY-----YIEPARGLDALREWIYNYTTERS 67
DB 26 GYFSAVYIEWVROAFKGLWVGEILPGSSNNSRYNE-KFKGVYVTRDTSTNTAYMELS 84

OY 68 KRDRRRRSVCARTWFCFRKYDVYRRIWHDITNTIS 106

DB 85 SLRSE-----DTAVYICARSTDFANFAWGGGTLVYS 117

RESULT 4

US-08-996-441B-109
Sequence 109, Application US/08996441B
Patent No. 6023013

GENERAL INFORMATION:

APPLICANT: English, Leigh H.

APPLICANT: Brussock, Susan M.

APPLICANT: Malvar, Thomas M.

APPLICANT: Bryson, James W.

APPLICANT: Kulesza, Caroline A.

APPLICANT: Walters, Frederick S.

APPLICANT: Slatin, Stephen L.

APPLICANT: Von Terssch, Michael A.

APPLICANT: Romano, Charles

TITLE OF INVENTION: INSECT-RESISTANT TRANSGENIC PLANTS

NUMBER OF SEQUENCES: 113

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Arnold, White & Durkee

STREET: P.O. Box 4433

CITY: Houston

STATE: Texas

COUNTRY: USA

ZIP: 77210

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/996,441B

FILING DATE: 18-DEC-1997

CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:

NAME: Kitchell, Barbara S.

REGISTRATION NUMBER: 33,928

REFERENCE/DOCKET NUMBER: MECO:151

TELECOMMUNICATION INFORMATION:

TELEPHONE: 512/418-3000

TELEPHONE: 512/474-7577

INFORMATION FOR SEQ ID NO: 109:

SEQUENCE CHARACTERISTICS:

LENGTH: 649 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-996-441B-109

Query Match 9.5%; Score 59; DB 3; Length 649;
Best Local Similarity 22.0%; Pred. No. 41;
Matches 22; Conservative 12; Mismatches 32; Indels 34; Gaps 4;

DB 239 NHCARW---YKGLDKRSTYE-----VRSIMHDTTNTISVS 75
16 NECVGLGVAVSANDVIRAAHGVYIEPEARGLDALREMIYNYTTERSKRRRR 75

DB 76 SVCHARTFCFRKYD-----VRSIMHDTTNTISVS 109
271 TLVVDLITFPLVDVRYTKGVTELTDRVLDPIYAVN 310

RESULT 5

US-08-483-101-14
Sequence 14, Application US/08483101
Patent No. 5932715

GENERAL INFORMATION:

APPLICANT: Scott, June R.
APPLICANT: Caron, Judy
TITLE OF INVENTION: CS2 Proteins and Coding Sequences
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Greenlee and Winner, P.C.
STREET: 5370 Manhattan Circle, Suite 201
CITY: Boulder
STATE: Colorado
COUNTRY: US
ZIP: 80303

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,101
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Farber, Donna M.
REGISTRATION NUMBER: 33878
REFERENCE/DOCKET NUMBER: 6-95
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 499-8080
TELEFAX: (303) 499-8089
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 890 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein
HYPOTHEICAL: NO

US-08-483-101-14

Query Match 9.4%; Score 58.5; DB 2; Length 890;
Best Local Similarity 31.0%; Pred. No. 71;
Matches 18; Conservative 10; Mismatches 21; Indels 9; Gaps 3;

DB 19 VGMGLV--AVSANDV-----IRAAHGVYIEPEARGLDALREMIYNYTTERSKR 69
488 MGVGCGESYASLSIPKGNLSLAYSNT--STSVRYDAVSEYVYVYVYVYVYVYVYVYV 543

RESULT 6

US-08-860-174A-4
Sequence 4, Application US/08860174A
Patent No. 5989830

GENERAL INFORMATION:

APPLICANT: DAVIS, Paul James
APPLICANT: VAN DER LOGT, Cornelis Paul Erik
APPLICANT: VERHOEIJEN, Martine Elisa
APPLICANT: WILSON, Steve
TITLE OF INVENTION: A BIFUNCTIONAL OR BIVALENT
TITLE OF INVENTION: ANTIBODY FRAGMENT ANALOGUE
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: PILLSBURY, MADISON & SUTRO LLP
STREET: 1100 New York Avenue, N.W.
STREET: 9th Floor, East Tower
CITY: WASHINGTON, D.C.
STATE:
COUNTRY: UNITED STATES
ZIP: 20005-3918

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS DOS Text
SOFTWARE: MS WORD, WINDOWS 95, Ver. 4.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/860,174A
FILING DATE: June 16, 1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 95307332.7
FILING DATE: October 16, 1995
APPLICATION NUMBER: PCT/EP/96/03605
FILING DATE: August 14, 1996

INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 135 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-860-174A-4

Query Match 9.2%; Score 57; DB 2; Length 135;
Best Local Similarity 25.0%; Pred. No. 9.2;
Matches 18; Conservative 10; Mismatches 18; Indels 26; Gaps 4;

DB 56 EMYYNY-----YTERSKRR---DRRRSVCH-----ARTWCFRKYDYVR 94

DB 68 EMGYITPYNDGTRKNEKATLTSKSSSTAYMELSLTSEDSAVYCSRRFY--- 125

DB 95 SIMHDTTNTIS 106
125 --WGGTTTVS 134

RESULT 7

US-08-659-251-8
Sequence 8, Application US/08659251
Patent No. 5883081

GENERAL INFORMATION:

APPLICANT: Kraus, Guenter
APPLICANT: Wong-Staal, Flossie
APPLICANT: Talbot, Randy
APPLICANT: Poeschla, Eric
TITLE OF INVENTION: Isolation of No. 5883081el HIV-2 Proviruses
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834

RESULT 8
PCT-US96-11445-8
Sequence 8, Application PC/TUS9611445
GENERAL INFORMATION:
APPLICANT: The Regents of the University of California
TITLE OF INVENTION: Isolation of Novel HIV-2 Proviruses
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: Robbins, Berliner & Carson
STREET: 201 N. Figueroa Street, 5th Floor
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90012-2628
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/11445
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Berliner, Robert
REGISTRATION NUMBER: 20,121
REFERENCE/DOCKET NUMBER: 3535-399C1

RESULT 9
 US-08-860-174A-12
 : Sequence 12, Application US/08860174A
 Patent No. 5989830
 GENERAL INFORMATION:
 APPLICANT: DAVIS, Paul James
 APPLICANT: VAN DER LOGT, Cornelis Paul Erik
 APPLICANT: VERHOEIJEN, Martine Elisa
 APPLICANT: WILSON, Steve
 TITLE OF INVENTION: A BIFUNCTIONAL OR BIVALENT
 TITLE OF INVENTION: ANIBODY FRAGMENT ANALOGUE
 NUMBER OF INVENTIONS: 31
 NUMBER OF SEQUENCES: 31
 CORRESPONDENCE ADDRESSES:
 ADDRESSEE: PILLSBURY, MADISON & SUTRO LLP
 STREET: 1100 New York Avenue, N.W.
 STREET: 9th Floor, East Tower
 CITY: WASHINGTON, D.C.
 STATE:
 COUNTRY: UNITED STATES
 ZIP: 20005-3918
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS DOS Text
 SOFTWARE: MS WORD, WINDOWS 95, Ver. 4.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/860,174A
 FILING DATE: June 16, 1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: EP 95307332.7
 FILING DATE: October 16, 1995
 APPLICATION NUMBER: PCT/EP/96/03605
 FILING DATE: August 14, 1996
 INFORMATION FOR SEQ ID NO: 12:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 274 amino acids
 type: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-860-174A-12

Best Local Similarity 25.0%; Pred. No. 23;
Matches 18; Conservative 10; Mismatches 18; Indels 26; Gaps 4;

QY 56 EWYIYV-----YERSRR-----DRRRSICH-----ARTWCFKRYVRR 94
Db 207 EWGYYTPYNDGKRYNEKFKATLTSDKSSYAMELSLTSEDSKVIYCSRRFDY--- 264

QY 95 SIWHTTTNTIS 106
Db 264 --WGQGTIVYS 273

RESULT 10
US-08-296-014A-4

Sequence 4, Application US/08296014A
Patent No. 5716834

GENERAL INFORMATION:

APPLICANT: Ding, Jeak Ling
TITLE OF INVENTION: The Cloned Factor C cDNA of the
TITLE OF INVENTION: Singapore Horseshoe Crab, Carcinoscoprius
TITLE OF INVENTION: roundicauda and Purification of Factor C Proenzyme
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch & Birch
STREET: 8110 Gatehouse Road, Suite 500 East
CITY: Falls Church
STATE: Virginia
COUNTRY: USA
ZIP: 22042

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/296,014A
FILING DATE:

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Murphy, Jr., Gerald M.
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 1781-105P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
TELEX: 248345
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1019 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-296-014A-4

Query Match 9.2%; Score 57; DB 1; Length 1019;
Best Local Similarity 23.2%; Pred. No. 1,3e+02;
Matches 23; Conservative 18; Mismatches 40; Indels 18; Gaps 5;

QY 21 WLGVAYSADVIVIRAAHEGYIEPEARGLDALREWIYNY-TERSKRRDRRRSVCH 79
Db 362 WGTAIYHELSSVCRAAIHAKL--PNSGAVHYVNGPYSDFLGSLNGIKSELSL-- 418

QY 80 ARTWCFKRYDYVRSI-----WHDTTTNTISYS 109
Db 418 ARSF---RFDYVRSYTAGKSGCPDGFVEVDENCVYVS 452

RESULT 11
US-08-596-405-4
Sequence 4, Application US/08596405
Patent No. 5858706

GENERAL INFORMATION:
APPLICANT: Ding, Jeak Ling
TITLE OF INVENTION: The Cloned Factor C cDNA of the
TITLE OF INVENTION: Singapore Horseshoe Crab, Carcinoscoprius
TITLE OF INVENTION: roundicauda and Purification of Factor C Proenzyme
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch & Birch
STREET: 8110 Gatehouse Road, Suite 500 East
CITY: Falls Church
STATE: Virginia
COUNTRY: USA
ZIP: 22042

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/596,405
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Murphy, Jr., Gerald M.
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 1781-105P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
TELEX: 248345
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1019 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-596-405-4

Query Match 9.2%; Score 57; DB 2; Length 1019;
Best Local Similarity 23.2%; Pred. No. 1,3e+02;
Matches 23; Conservative 18; Mismatches 40; Indels 18; Gaps 5;

QY 21 WLGVAYSADVIVIRAAHEGYIEPEARGLDALREWIYNY-TERSKRRDRRRSVCH 79
Db 362 WGTAIYHELSSVCRAAIHAKL--PNSGAVHYVNGPYSDFLGSLNGIKSELSL-- 418

QY 80 ARTWCFKRYDYVRSI-----WHDTTTNTISYS 109
Db 418 ARSF---RFDYVRSYTAGKSGCPDGFVEVDENCVYVS 452

RESULT 12
US-08-877-620-4
Sequence 4, Application US/08877620
Patent No. 5985590

GENERAL INFORMATION:
APPLICANT: Ding, Jeak Ling
TITLE OF INVENTION: The Cloned Factor C cDNA of the
TITLE OF INVENTION: Singapore Horseshoe Crab, Carcinoscoprius
TITLE OF INVENTION: roundicauda and Purification of Factor C Proenzyme
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch & Birch
STREET: 8110 Gatehouse Road, Suite 500 East
CITY: Falls Church
STATE: Virginia
COUNTRY: USA
ZIP: 22042
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/877,620
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/596,405
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Murphy, Jr., Gerald M.
 REGISTRATION NUMBER: 28,977
 REFERENCE/DOCKET NUMBER: 1781-105P
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703) 205-8000
 TELEFAX: (703) 205-8050
 TELETYPE: 248345
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1019 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-877-620-4

Query Match 9.2%; Score 57; DB 2; Length 1019;
 Best Local Similarity 23.2%; Pred. No. 1.3e+02;
 Matches 23; Conservative 18; Mismatches 40; Indels 18; Gaps 5;
 QY 21 WGVASAVVDVIRAAHEGYIEPANGRLDALREMIYNTY-TERSKRRDRRRSYCH 79
 DB 362 WGTATYHEHSYCRALHAGKL--PNSGAVHYVNNNGPYSDPLGSDLNIGKSEELKSL-- 418
 QY 80 ARTWCFRRKYDVRSI-----WHTTNTISYVS 109
 DB 418 ARSF-----RFDYVRSTKSGSCDPCGFVDENCVYVTS 452

RESULT 13
 US-08-471-119A-2
 Sequence 2, Application US/08471119A
 Patent No. 5827706
 GENERAL INFORMATION:
 APPLICANT: Leitner, Ernst
 APPLICANT: Schneider, Elisabeth
 APPLICANT: Schoergerdorfer, Kurt
 APPLICANT: Weber, Gerhard
 TITLE OF INVENTION: Cyclosporin Synthetase
 NUMBER OF SEQUENCES: 8
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: No. 5827706artis Corporation
 STREET: 59 Route 10
 CITY: East Hanover
 STATE: New Jersey
 COUNTRY: USA
 ZIP: 07936
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/471,119A
 FILING DATE: 06-JUN-1995
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Kassenoft, Melvyn
 REGISTRATION NUMBER: 26,389
 REFERENCE/DOCKET NUMBER: 100-8029/CONT/CONT
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 201 503 8474

TELEFAX: 201 503 8807
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 15281 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: unknown
 MOLECULE TYPE: protein
 HYPOTHETICAL: NO
 AMTI-SENSE: NO
 ORIGINAL SOURCE:
 ORGANISM: Tolypocladium niveum
 STRAIN: ATCC 34921
 US-08-471-119A-2

Query Match 9.2%; Score 57; DB 2; Length 15281;
 Best Local Similarity 30.5%; Pred. No. 4.5e+03;
 Matches 18; Conservative 7; Mismatches 22; Indels 12; Gaps 2;
 QY 11 VCDs-----ONECVGLGVASAVVDVIRAAHEGYIEPEA--RGRLDALREW 57
 DB 3844 ICESYDLGGDAKSDNDYRWSLSARSNVAVASISALDVIDIAQEAQFVEISCAROW 3902

RESULT 14
 US-08-379-057-31
 Sequence 31, Application US/08379057
 Patent No. 5876950
 GENERAL INFORMATION:
 APPLICANT: Sladak, Anthony W.
 APPLICANT: Hollenbaugh, Diane L.
 APPLICANT: Gilliland, Lisa K.
 APPLICANT: Gordon, Marcia L.
 APPLICANT: Bajorath, Jurgen
 APPLICANT: Aruffo, Alejandro A.
 TITLE OF INVENTION: Monoclonal Antibodies Specific For
 TITLE OF INVENTION: Different Epitopes of Human gp99 and Methods For Their Use
 NUMBER OF SEQUENCES: 57
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Bristol-Myers Squibb Company
 STREET: 3005 First Avenue
 CITY: Seattle
 STATE: Washington
 COUNTRY: USA
 ZIP: 98121
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/379,057
 FILING DATE: 26-JAN-1995
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Poor, Brian W.
 REGISTRATION NUMBER: 32,928
 REFERENCE/DOCKET NUMBER: ON0133-
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 727-3670
 TELEFAX: (206) 727-3601
 INFORMATION FOR SEQ ID NO: 31:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 115 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 FRAGMENT TYPE: internal
 US-08-379-057-31

Mon Jun 26 09:10:58 2000

us-09-214-478-4.rai

Page 8

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 23, 2000, 09:17:09 ; Search time 592.38 Seconds
(without alignments)
2360.583 Million cell updates/sec

Title: US-09-214-478-3

Sequence: 345
1 atggtcttcacgctcttc.....cggcgactcgcgtaacgtag 345

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4857316 segs, 2026611650 residues

Total number of hits satisfying chosen parameters: 9714632

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : EST:
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*

45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_est45:*
80: gb_est46:*
81: gb_est47:*
82: gb_gsa1:*
83: gb_gsa2:*
84: gb_gsa3:*
85: gb_gsa4:*
86: em_gsa1:*
87: em_gsa2:*
88: em_gsa3:*
89: em_gsa4:*
90: gb_gsa5:*
91: gb_gsa6:*
92: gb_gsa7:*
93: gb_gsa8:*
94: gb_gsa9:*
95: em_gsa5:*
96: em_gsa6:*
97: em_gsa7:*
98: em_gsa8:*
99: em_gsa9:*
100: em_gsa10:*
101: em_gsa11:*
102: gb_gsa10:*
103: gb_gsa11:*
104: em_gsa12:*
105: gb_gsa12:*
106: gb_gsa13:*
107: gb_gsa14:*
108: gb_gsa15:*
109: gb_gsa16:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result

Query

No.	Score	Match length	DB	ID	Description
1	303	87.8	630	79	AM276836
2	138.4	40.1	426	79	AM265066
3	35.4	10.3	503	26	W18889
4	35.4	10.3	345	35	AA656037
5	35.4	10.3	348	35	AA656203
6	35.4	10.3	389	34	AA659726
7	35.4	10.3	453	26	W30475
8	35.4	10.3	477	27	AA041778
9	35.4	10.3	479	27	W82078
10	35.4	10.3	504	39	AA874208
11	35.4	10.3	543	60	AI188504
12	35.4	10.3	543	29	AA140104
13	35.4	10.3	547	31	AA899263
14	35.4	10.3	551	60	AI768495
15	35.4	10.3	556	43	AI195511
16	35.4	10.3	565	30	AA217816
17	35.4	10.3	655	47	AI481623
18	35.4	10.3	976	42	AI115246
19	35.4	10.3	945	26	W41762
20	35	10.1	684	44	AI322378
21	34.8	10.1	581	102	AO162258
22	34.8	10.1	696	105	AO335183
23	34.8	10.1	803	91	AO891726
24	34	9.9	697	69	AM107119
25	33.8	9.8	312	43	AI195796
26	33.8	9.8	339	81	AM412524
27	33.8	9.8	489	33	AA404076
28	33.8	9.8	662	28	AA123015
29	33.6	9.7	530	63	AI199408
30	33.4	9.7	909	82	CNS00507
31	32.8	9.5	661	109	AO658467
32	32.2	9.3	483	27	W82481
33	32.2	9.3	545	25	AI1885950
34	31.4	9.1	241	32	AA377776
35	31.4	9.1	748	82	AA001595
36	31.2	9.0	449	63	AM006018
37	31.2	9.0	659	72	AM163374
38	31.2	9.0	710	92	AO944247
39	31	9.0	256	25	W06958
40	31	9.0	360	91	AO841382
41	31	9.0	407	91	AO841373
42	30.8	8.9	465	22	AI18484
43	30.8	8.9	739	82	AG001599
44	30.6	8.9	354	64	AM038817
45	30.6	8.9	752	37	AA703929

IGNITIONS

RESULT 1
AM276836 LOCUS
DEFINITION xp66g11.xl NCI_CGAP bp EST 03-JAN-2000
similar to SW:441_ADE02 P03240 PROBABLE EARLY E4 I3 KD PROTEIN.
mRNA sequence.
ACCESSION AM276836
VERSION AM276836
KEYWORDS GI:6663866
SOURCE EST.
ORGANISM human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo. "
1 (bases 1 to 630)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
On Nov 22, 1999 this sequence version replaced gi:6662296.
Contact: Robert Strausberg, Ph.D.
Tel.: (301) 496-1550

Email: Robert.Strausberg@nih.gov
Tissue Procurement: Andrew Bercchuck M.D., John Gillespie M.D.,
Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: David B. Krizman, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E Consortium/GENB at:
www-bio.lnl.gov/bbrp/image/image.html

Possible reversed clone: similarity on wrong strand
Possible reversed clone: polyT not found
Seq primer: -40UP from Gibco
High quality sequence stop: 422.

FEATURES

SOURCE

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2745380"
/clone_11b="NCI_CGAP_OV39"
/sex="Female"
/tissue_type="papillary serous ovarian metastasis"
/lab_host="DH103"
/notes="Organ: ovary; Vector: PAMP10; cDNA made by oligo-dT priming. Non-directionally cloned into the XbaI sites of PAMP10. Size-selected on agarose gel, average insert size 500 bp. Primary library: non-amplified. cDNA library preparation: David B. Kriizman, Ph.D (NCI). Reference: Kriizman et al. (1996) Cancer Research 56:15380-15383."

```

[illegible]

Query Match	87.8%	Score 303;	DB 79;	Length 630;
Best Local Similarity	98.1%	Pred. No. 1.5e-85;		
Matches 306; Conservative	0;	Mismatches 6;	Indels 0;	Gaps 0;

QY 1 atggttcctcagctctcccgctcctccggtgtgtgactcgcagaacgaatgttagt 60

Db 319 ATGGTTCCTCCAGCTCTTCCCGCTCCTCCCGTGTGTGACTCGCAGACGAATGTGAGGT 378

61 tggctggtgtggccttattctgcggtggtgatatgtatcagggcagcgcgcataaaga 120

Db 379 TGGCTGGGTGTGGCTTATCTCGGGTGGTGGATGTTATCANGCAGCGGCCATGAAGA 438

QY 121 gttacatagaaccggaagccaggggcgccctgatatgtcttgagagagtgtgatatatactac 180

Db 439 GTTACATAGAACCCGAAGCCAGGGGGCGCCTGGATGCTTGAAGAGAGTGATATACTAC 498

QY 181 aactactacagagcgtctaagcgcgagaccggagacgcagatctgtgtcacgcc 240

Db 499 AACTACTACACAGAGCGAGCTAAGCGACGAGACCGGAGAGCAGATCTGTTGTCACGCC 558

QY 241 cgcacctgttttgcctcaggaatatgactacgtccggcgttccatttggcatgacct 300

Db 559 CGCACCCTGGTTTTCCTTCAAGAAATATGACTACGTGGGGCGTTGCATTTGGCATGACACT 618

QY 301 acgaccaacacg 312

D_b 619 ACGACCAACACG 630

РЕШЕНИЕ 2 1 1 9

AW265066	AW265065	435 ps	EBVT	ECT	28-DEC-1999
LOCUS					

DEFINITION	NCI_CGAP_Co22	Homo sapiens	CDNA clone	IMAGE:2755060 3'
xq60g03.x1	SW:E411	AB03	AB03241	PROBABLE EXPR E4 11 3D PROMETN

ACCESSION [1] ;, mRNA sequence.
AW365066

```

KEYWORDS
VERSION AW265066.1
EST      GI:6641882

```

SOURCE	human.	..5.
--------	--------	------

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 426)
AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jul 7, 1999 this sequence version replaced gi:5406301.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Unknown library type
Possible reversed clone: similarity on wrong strand
Possible reversed clone: polyt not found
Seq primer: -400p from Glibco
High quality sequence stop: 425.
Location/Qualifiers
1. 426
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2755060"
/clone_lib="NCI-CCAP-Co22"
/issue_type="Colon adenocarcinoma"
/lab_host="DH10B"
/note="Organ: Colon; Vector: PAMP10; CDNA made by oligo-
priming. Non-directionally cloned into the UDG sites of
PAMP10. Size-selected on agarose gel, average insert
size 500 bp. Primary library: non-amplified. CDNA
Library Preparation: David B. Kitzman, Ph.D. (NCI).
Reference: Kitzman et al. (1996) Cancer Research
56:5380-5383."

BASE COUNT 103 a 80 c 118 g 124 t 1 others
ORIGIN

Query Match 40.1%; Score 138.4; DB 79; Length 426;
Best Local Similarity 98.6%; Pred. No. 1.7e-33;
Matches 139; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 aggtgtttccagctcttcctccgtctccgtgtgtgactccagagcaatgttagt 60
DB 286 ATGGTCTTCCAGCTCTTCCGCTCTCCGCTGTGACTCGCAGAACAAATGTAGGT 345
QY 61 tggctgtgtgtgcttattctgcgtgtgtgtgtatcaaggcagcgatgaagga 120
DB 346 TGGCTGTGTGTGCTTATTCTGCGGTGTGTGATGATCAAGGACGCGGCGATGAAGA 405
QY 121 gttacatgaaccggaagcc 141
DB 406 GTATACATGAACCCGAAC 426

RESULT 3
W18889 303 bp mRNA EST 10-SEP-1996
LOCUS W18889
DEFINITION mc04g02.r1 Soares mouse p33NF19.5 Mus musculus cDNA clone
IMAGE:337970 5' similar to SW:RB14_RAT P35287 RAS-RELATED PROTEIN
RAB-14. [1], mRNA sequence.
W18889
ACCESSION W18889.1 GI:1294599
VERSION EST.
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 303)
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
The WashU-HMI Mouse EST Project

JOURNAL Unpublished (1996)
COMMENT On May 9, 1995 this sequence version replaced gi:802677.
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
NCI:219370
Possible reversed clone: similarity on wrong strand
Seq primer: mod.RBGA+RT
High quality sequence stop: 274.
Location/Qualifiers
1. 303
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:337970"
/clone_lib="Soares mouse p33NF19.5"
/dev_stage="19.5 dpc total fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Vector: pT73D (Pharmacia) with a modified
polylinker. Site 1: Not I. Site 2: Eco RI; 1st strand cDNA
was primed with a Not I...oligo(NT) primer [5',
TGTTACCAATCTGAAGTGGAGGCGCCGCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Benito
Soares and M. Fatima Bonaldo. RNA was kindly provided by
Dr. Minoru Ko (Wayne State University)."

BASE COUNT 106 a 61 c 77 g 59 t
ORIGIN

Query Match 10.3%; Score 35.4; DB 26; Length 303;
Best Local Similarity 52.3%; Pred. No. 0.6;
Matches 78; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 75 ttattctcgtgtgtgtgtatgtatagggcagcgcatgaaggtttacatagaacc 134
DB 8 TCAACTGCGAGTTTGGGATACAGCAGGACGACGGCTTACAGCCGTTACAGGACT 67
QY 135 cgaagccaggggagcgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 194
DB 68 ACTATAGAGGAGCTGACAGTGGCGTCATGCTATGACATCACCAGAAAGTACATATA 127
QY 195 ggcattcaagcgcgagacggagacgca 223
DB 128 ACCACTTAAGCAGCTGGTTGACACGCA 156

RESULT 4
AA656067 345 bp mRNA EST 04-NOV-1997
LOCUS AA656067
DEFINITION vs49h11.r1 Stratagene mouse Tcell 937311 Mus musculus cDNA clone
IMAGE:1149669 5' similar to SW:RB14_RAT P35287 RAS-RELATED PROTEIN
RAB-14. [1], mRNA sequence.
AA656067
ACCESSION AA656067.1 GI:2592221
VERSION EST.
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 345)
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.

TITLE The MasbU-HMT Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Sep 12, 1996 this sequence version replaced g1:1393029.
Contact: Mairra M/Mouse EST Project
MasbU-HMT Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu

Source

FEATURES
Source

This clone is available royalty-free through LNC ; contact the IMAGE Consortium (Image.lnci.gov) for further information.
 MGI:247220
 Possible reversed clone: similarly on wrong strand
 Seq primer: -28M13 rev2 from Amersham
 High quality sequence stop: 449.

Best Local

BASE COUNT
ORIGIN

BASE COUNT
ORIGIN

W82078
TODTS

27	455
Db	208

SOURCE :

house mouse.
Mus musculus

AUTHORS

Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi, Muridae; Murinae; Mus. 1 (bases 1 to 479)
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Gelsel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,

TITLE
JOURNAL
COMMENT

FEATURES
Source

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:403452"
/clone_1lb="Soares mouse embryo NDM13.5 14.5"
/sex="unknown"
/tissue_type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="DH10B"
/notes="vector: pT73b-Pac (Pharmacia) with a modified
polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5,
TGTTCACAACTCGAAGTGGGCGCGCGCGAATTTTTTTTTTTTTTTTTT
T 3' ], on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos [total RNA provided by Minoru Ko, Wayne
State Univ. from 2 ]; double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT73 vector. Library went through one round of
normalization, and was constructed by Bento Soares and
M.Fatima Bonaído. "

```

BASE COUNT
ORIGIN

Query Match	10.38;	Score 35.4;	DB 27;	Length 479;
Best Local Similarity	52.3%;	Pred. No. 0.74;		
Matches 78;	Conservative 0;	Mismatches 71;	Indels 0;	Gaps 0;
QY	75	ttattctgctgctgctgattgtatcagaagcgagcgccatgaaagagtttaacataaac	134	
Db	88	TCAAACTGCAGATTGGGATACACAGAGGCGAGCGCGTTCAACAGCGGTTACACGAGCT	147	
QY	135	cgaagccagagggcgccctgagatgcttggagaagatgatatactaacactacaaga	194	
Db	148	ACGATGAGGAGACTCTCAGATGCGCTCACTGGTGTATGATCATCCACCAAAAGATGAATATA	207	

RESULT 10

QY 195 gcatcttagcgcgagaccgagagca 223
 | | | | | | | | | | | | | | | |
 Db 208 ACCACTTAAGCAGCTGGTTGACAGAGCA 236

LOCUS

LOCUS	504 bp	MRNA	EST	19-MAR-1998
DEFINITION	AA874208			
	VX03g04.r1	Soares_thymus_2NDMT	Mus musculus	CDNA clone

DEFINITION

IMAGE:1263414 5' similar to SW:RB14_PAT P35287 RAS-RELATED PROTEIN
RAB-14. ; mRNA sequence.
11074000
CCGCCTCAT

KEYWORDS

KEYWORDS	EST.
SOURCE	house mouse


```

Db      408 TCAACTGCAGATTGGATACAGCAGGCGGTTACAGCGGTACACGGAGCT 467
Oy      135 cgaagccagggggggcgcttgatgagagagtgatataactactacacaga 194
Db      468 ACTATAGAGAGAGCTGCGGCTTCATGATGATACATCACCAAGAGAGTACATATA 527
Oy      195 gcgatcctaagcgagagacggagagcga 223
Db      528 ACCACTTAAGCAGCTGGTTGACAGAGCGCA 556

```

Search completed: June 23, 2000, 09:17:14
 Job time: 12882 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 23, 2000, 10:05:42 ; Search time 56.99 seconds
(without alignments)
1514.586 Million cell updates/sec

Title: US-09-214-478-3

Perfect score: 345
Sequence: 1 atgtgtcttcacagctcttc.....cggcgcactccgctacagtag 345

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: N_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match length	ID	Description
1	345	100.0	1 T31315	Adenovirus E4 and Complete sequence
2	345	100.0	1 V232370	Complete sequence
3	345	100.0	1 V232375	Complete sequence
4	345	100.0	1 T60559	Recombinant adenovirus
5	345	100.0	1 V07261	Adenoviral vector
6	345	100.0	1 X15627	Recombinant adenovirus
7	345	100.0	1 T60557	Recombinant cis-ec
8	345	100.0	1 V07258	Adenovirus 5 genome
9	345	100.0	1 T60558	Recombinant trans-Adenovirus vector
10	343.4	99.5	1 X07371	Adenovirus vector
11	143.8	41.7	1 V21141	Chimpanzee adenovirus
12	103	29.9	1 V21140	Sequence encoding Ad2/ORF6/PGK-CFR
13	85.4	24.8	1 T58270	Plasmid pMTV4ORF
14	80	23.2	1 T58273	Second generation ChimERIC plasmid p
15	80	23.2	1 V21144	Glucocorticoid Rec
16	80	23.2	1 T58273	Adenovirus 17. Rec
17	78.8	22.8	1 T31320	Adenovirus 17. Rec
18	54	15.7	1 V21112	Human melanoma ass
19	30.8	8.9	1 T77840	Human melanoma ass
20	30.8	8.9	1 T77839	Human melanoma ass
21	30.8	8.9	1 T77838	Human melanoma ass
22	30.4	8.8	1 T38805	Mouse cortactin ge
23	29	8.4	1 T38805	Tobacco clip-53 NE
24	29	8.4	1 T38805	Sequence encoding
25	29	8.4	1 T38805	Polynucleotide enc
26	29	8.4	1 T38805	Sequence encoding
27	29	8.4	1 T38805	PART gene promoter
28	29	8.4	1 T38805	PRTc 99 plasmid fr
29	29	8.4	1 T38805	PART gene promoter
30	29	8.4	1 T38805	PUMGIR sequence 1
31	29	8.4	1 T38805	pDE4: Plasmid DNA
32	29	8.4	1 T38805	Plasmid pGFP190-1
33	29	8.4	1 T38805	Plasmid DGP232-4 C
34	29	8.4	1 T38805	Sequence of pTET4.

35	29	8.4	7742	1 V03430	Escherichia coli g
36	29	8.4	9299	1 V63730	Vector plasmid per
37	29	8.4	9335	1 V63734	Plasmid pGP/367 c
38	29	8.4	9408	1 V63731	Vector plasmid PER
39	29	8.4	10160	1 V63723	Vector plasmid PDA
40	29	8.4	11784	1 V63724	Vector plasmid PDA
41	29	8.4	11991	1 V63725	Vector plasmid PDA
42	29	8.4	14698	1 V61792	Marker plasmid Des
43	29	8.4	19001	1 V61793	Target plasmid M01
44	29	8.4	19035	1 V61794	Target plasmid Man
45	28.2	8.2	3060	1 Q03098	Enod2a genomic clo

ALIGNMENTS

RESULT 1	1	100.0%	Score 345; DB 1; Length 3189;
ID T31315/c	1	100.0%	Score 345; DB 1; Length 3189;
AC T31315;	1	100.0%	Score 345; DB 1; Length 3189;
DT 16-APR-1997 (first entry)	1	100.0%	Score 345; DB 1; Length 3189;
DE Adenovirus E4 and part ITR region corresp. to bases 32749-35935.	1	100.0%	Score 345; DB 1; Length 3189;
KW Human adenovirus; inverted tandem repeat; E4; E1; open reading frame;	1	100.0%	Score 345; DB 1; Length 3189;
KW expression; late gene; cell line; defective recombinant adenovirus;	1	100.0%	Score 345; DB 1; Length 3189;
KW complementation; gene therapy; cystic fibrosis; dystrophy; ds;	1	100.0%	Score 345; DB 1; Length 3189;
OS neurodegenerative disease; blood coagulation disorder; viral infection.	1	100.0%	Score 345; DB 1; Length 3189;
OS Adenovirus Ad5.	1	100.0%	Score 345; DB 1; Length 3189;
PN MO9622378-A1.	1	100.0%	Score 345; DB 1; Length 3189;
PD 25-JUN-1996	1	100.0%	Score 345; DB 1; Length 3189;
PR 19-JAN-1996; F00088	1	100.0%	Score 345; DB 1; Length 3189;
PR 20-JAN-1995; FR-000747.	1	100.0%	Score 345; DB 1; Length 3189;
PR 01-JUN-1995; FR-006532.	1	100.0%	Score 345; DB 1; Length 3189;
PR 08-SEP-1995; FR-010541.	1	100.0%	Score 345; DB 1; Length 3189;
PA (RHON) RHONE POTLENC RORER SA.	1	100.0%	Score 345; DB 1; Length 3189;
PI Dedieu J, Latta M, Orsini C, Perricaudet M, Vayne E;	1	100.0%	Score 345; DB 1; Length 3189;
PI Yeh P, Dedieu JF;	1	100.0%	Score 345; DB 1; Length 3189;
DR WPI: 96-35433/35.	1	100.0%	Score 345; DB 1; Length 3189;
PT Cells for prodn. of recombinant adeno and adeno-associated virus	1	100.0%	Score 345; DB 1; Length 3189;
PT contain part of the E4 region of adeno virus, used to produce new	1	100.0%	Score 345; DB 1; Length 3189;
PT adeno virus defective in E1 and E4 regions for use as gene therapy	1	100.0%	Score 345; DB 1; Length 3189;
PT vectors	1	100.0%	Score 345; DB 1; Length 3189;
PS Disclosure: Page 37-39; 69pp; French.	1	100.0%	Score 345; DB 1; Length 3189;
CC This is the sequence corresponding to nucleotides 32749-35935 of the	1	100.0%	Score 345; DB 1; Length 3189;
CC human adenovirus type 5 (Ad5), which covers the E4 and right inverted	1	100.0%	Score 345; DB 1; Length 3189;
CC tandem repeat (ITR) region of the viral genome. The E4 region of the	1	100.0%	Score 345; DB 1; Length 3189;
CC genome contains 7 open reading frames (ORFs 1-7). These ORFs are thought	1	100.0%	Score 345; DB 1; Length 3189;
CC to regulate expression of the adenoviral late genes. The E4 region is	1	100.0%	Score 345; DB 1; Length 3189;
CC used to generate cell lines, esp. 293 cells, which can support the	1	100.0%	Score 345; DB 1; Length 3189;
CC production of defective recombinant adenoviruses, esp. those that lack	1	100.0%	Score 345; DB 1; Length 3189;
CC the E1 and part of the E4 region. The cell lines pref. have the E4	1	100.0%	Score 345; DB 1; Length 3189;
CC region inserted into their genomes for complementation of the defective	1	100.0%	Score 345; DB 1; Length 3189;
CC viruses. The E4 region is inserted under control of a promoter pref.	1	100.0%	Score 345; DB 1; Length 3189;
CC from MMTV (Pharmacia) which is inducible by dexamethasone. Four defective	1	100.0%	Score 345; DB 1; Length 3189;
CC recombinant viruses are esp. mentioned in the specification: they all	1	100.0%	Score 345; DB 1; Length 3189;
CC lack all or part of the E1 region and part of the E4 region; virus AV1	1	100.0%	Score 345; DB 1; Length 3189;
CC also lacks nucleotides (nts) 34801-34329 and 34115-33126 consisting of	1	100.0%	Score 345; DB 1; Length 3189;
CC ORF3 and 6. AV2 also lacks all of the E4 region except for ORF1 (nts	1	100.0%	Score 345; DB 1; Length 3189;
CC 33093-35053). AV3 also lacks all of the E4 region except for ORF4	1	100.0%	Score 345; DB 1; Length 3189;
CC (present as nts 33093-33695 and 34634-33355). AV4 further lacks the	1	100.0%	Score 345; DB 1; Length 3189;
CC ORF5 of the E4 region i.e. nts 32720-35835, 33466-35355 or 33093-35355.	1	100.0%	Score 345; DB 1; Length 3189;
CC The defective viruses generated in the new cells can be used for gene	1	100.0%	Score 345; DB 1; Length 3189;
CC therapy e.g. for cystic fibrosis, dystrophy, neurodegenerative diseases,	1	100.0%	Score 345; DB 1; Length 3189;
CC blood coagulation disorders, viral infections, etc.	1	100.0%	Score 345; DB 1; Length 3189;
CC Sequence 3189 BP; 975 A; 912 C; 631 G; 671 T;	1	100.0%	Score 345; DB 1; Length 3189;
SO	1	100.0%	Score 345; DB 1; Length 3189;
Query Match	1	100.0%	Score 345; DB 1; Length 3189;
Best Local Similarity	1	100.0%	Score 345; DB 1; Length 3189;
Matches 345; Conservative	1	100.0%	Score 345; DB 1; Length 3189;
DB 1593 ATGTTCTTCACAGCTCTTCCTCCGCTGCTGACTCCAGACAGAAATGTAGGT 1534	1	100.0%	Score 345; DB 1; Length 3189;

QY	61	tgctcgggtgtgctgattatcttcgcggttggtgtgattcttctacgaagcgagcgcgccatgaaga	120
Db	1533	tgcctgggtgtgctgattatcttcgcggttggtgtgattcttctacgaagcgagcgcgccatgaaga	1474
QY	121	gtttacatagaaccgccgaagccagcgagcgcgccctgatagtgttttgaagagtgtatactac	180
Db	1473	gtttacatagaaccgccgaagccagcgagcgcgccctgatagtgttttgaagagtgtatactac	1414
QY	181	aactctcacagagcagttcttaagcgcgagcgagacccgaagcagaatctcgtttgtcaagcc	240
Db	1413	aactctcacagagcagttcttaagcgcgagcgagacccgaagcagaatctcgtttgtcaagcc	1354
QY	241	cgcacctggtttgttcttaaggaatactactacgtcgcgcgcttcatttgcatacact	300
Db	1353	cgcacctggtttgttcttaaggaatactactacgtcgcgcgcttcatttgcatacact	1294
QY	301	acgacacaacagatctcggtgtgtctcgcgcgacatccgtacagtag	345
Db	1293	acgacacaacagatctcggtgtgtctcgcgcgacatccgtacagtag	1249
RESULT 2			
ID	V33370/C		
AC	V33370 standard; DNA: 8710 BP.		
DT	23-SEP-1998 (first entry)		
DE	Complete sequence of the pE4/Hygro plasmid.		
KE	Circular; adenovirus type 5; pE4/Hygro plasmid; structural protein;		
KW	Complementation; E4 regulatory protein, gene therapy; HIV; tumour;		
KW	Huntington's disease; Tay-Sachs disease; sickle cell disease;		
KW	early gene; ds.		
OS	Synthetic.		
FH	Key	Location/Qualifiers	
FT	misc_feature	complement (707..3820)	
FT		/*tag= a	
FT		/note= "E4 regulatory gene"	
FT		3830..6470	
FT	misc_feature	/*tag= b	
FT		/note= "Hygromycin resistance gene"	
PN	W09813499-A2.		
PD	02-APR-1998.		
PE	24-SEP-1997; E05251.		
PR	25-SEP-1986; US-719806.		
PA	(NOVS) NOVARTIS AG.		
PA	(SCRI) SCRIPS RES INST.		
PI	Memrow GR, Von Seggern DJ;		
DR	WPI; 98-230709/20.		
PT	Adenoviral vectors - which lack DNA encoding for structural protein		
PT	or fibre protein used particularly for gene therapy		
PS	Example 1; Pages 63-74, 170pp; English.		
CC	The present sequence is that of a pE4/Hygro plasmid used in the		
CC	method of the invention. The plasmid contains an adenovirus type 5		
CC	E4 regulatory gene and a hygromycin resistant gene. The invention		
CC	provides adenoviral vectors having deletions of all or part of		
CC	various gene sequences encoding adenoviral structural proteins.		
CC	Deletions in the structural proteins would allow a reduced risk of		
CC	wild-type virus contamination and would also allow packaging of foreign		
CC	DNA in such vectors for a variety of diagnostic and therapeutic		
CC	applications. The adenoviral vectors having deletions in the structural		
CC	gene regions are produced by cellular complementation of these		
CC	adenoviral genes. Therefore, the pE4/Hygro plasmid was used as a		
CC	complementation plasmid which was introduced into a host cell line		
CC	where parts of the E4 gene region would be stably inserted into the		
CC	host cell chromosomes. The resulting E4 deficient plasmid can be used		
CC	as a gene delivery vector. The vectors can be used for diagnosis or		
CC	gene therapy, e.g. for treating conditions characterised by		
CC	hyper-proliferative cells (e.g. tumours), genetic diseases (e.g.		
CC	Huntington's disease, Tay-Sachs disease), or sickle cell disease), or		
CC	infections (e.g. HIV infection). They can also be used for in vitro		
CC	production of biologically active proteins.		
CC	Sequence 8710 BP; 2263 A; 2365 C; 2141 G; 1941 T;		

Query Match	Similarity	100.0%	Score 345	DB 1	Length 8710
Best Local	Similarity	100.0%	Pred. No. 8.2e-105		
Matches 345	Conservative	0	Mismatches	0	Indels
					Gaps 0
OY	1	atggtcttcacagcgccttcctccgctcctccgctggttgactcgcaagcaatggtgagt	60		
DB	2382	ATGGTCTTCCAGCCTCTTCCGCTCTCTCCCGTGTGACTCGCAGACGAATGTGTAGGT	2323		
OY	61	tggctggtgtgtggtctatctctgcgtgtgtgtatcattcaaggcagcgcatgaaga	120		
DB	2322	TGGCTGGGAGTGTGGCTTATCTCGGTGTGTGATGTATTACAGGCGACGCGCATGAGGA	2263		
OY	121	gtttacatagaaccccggaagccaggggggcccctgagctgtcttaagagatgagtataaac	180		
DB	2262	GTTTACATRGAACCCGGAAGCCAGGGGGCCCCCTGGATGCTTTAGAGAGGGATATRATC	2203		
OY	181	aactactacacagcagcatctaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	240		
DB	2202	AACCTACTACACAGCAGCATCTAAGGGGGGAGACCGGAGACGCGAGATCTGTTGTACGGC	2143		
OY	241	cgccactggtttgtcttaaggaataatactacgtccggcggttccattgtgcataaact	300		
DB	2142	CGCACCTGGTGTGCTTACGGAATAATGATCTACGTCCGCGGCTTCATTGGCATGACACT	2083		
OY	301	agggacaaacagcatctggtgtgtctcgcgcgcctccgtacaagtag	345		
DB	2082	ACGACCAACACGATCTCGGTGTCTCGGGCGACCTCGTACACTAG	2038		
RESULT	3				
V32375/c					
ID	V32375	standard; DNA; 10610 bp.			
AC	V32375				
DT	29-SEP-1998	(first entry).			
DE		Complete sequence of the pE4/fiber plasmid.			
KW		Circular; adenovirus type 5; pE4/fiber plasmid; structural protein;			
KW		complementation; fiber protein; gene therapy; HIV; tumour; AD5;			
KW		early gene; Huntington's disease; Tay-Sachs disease;			
KW		sickle cell disease; E4 regulatory gene; ds.			
HH		Synthetic.			
FT	Key	Location/Qualifiers			
FT	misc_feature	complement (21..3149)			
FT		/*tag= a			
FT	sig_peptide	/note= "AD5 E4 regulatory gene"			
FT		4051..4366			
FT		/*tag= b			
FT		/note= "AD5 leader sequence"			
FT		4372..6124			
FT	misc_feature	/*tag= c			
FT		/note= "AD5 fiber gene"			
PN	W09813499-A2.				
PD	02-APR-1998.				
PF	24-SEP-1997; E05251.				
PR	25-SEP-1996; U5-719806.				
PA	(NOVS) NOVARTIS AG.				
PA	(SCRI) SCRIPPS RES INST.				
PI	Memerow GR, Von Seggern DJ;				
PR	WPI; 98-230709/20.				
PT	Adenoviral vectors - which lack DNA encoding for structural protein				
PT	or fibre protein used particularly for gene therapy				
PS	Example 1; Pages 131-145; 170pp; English.				
CC	The present sequence is that of a pE4/fiber plasmid used in the				
CC	method of the invention. The plasmid contains an adenovirus type 5 (AD5)				
CC	fiber gene controlled by a CMV promoter, an AD5 E4 gene and an adenovirus				
CC	type 2 (AD2) tripartite leader sequence upstream of the fiber gene. The				
CC	invention provides adenoviral vectors having deletions of all or part of				
CC	various gene sequences encoding adenoviral structural proteins and/or				
CC	early region proteins. Deletions in these proteins would allow a				
CC	reduced risk of wild-type virus contamination and would also allow				
CC	packaging of foreign DNA in such vectors for a variety of diagnostic and				
CC	therapeutic applications. The adenoviral vectors having deletions in				
CC	the structural and/or early gene regions are produced by cellular				

CC complementation of these adenoviral genes. Therefore, the pB4/Fiber
CC plasmid was used as a complementation plasmid which was introduced into
CC a host cell line where parts of the fiber and E4 gene region would be
CC stably inserted into the host cell chromosomes. The resulting E4/fiber
CC gene deficient plasmid can be used as a gene delivery vector. The
CC vectors can be used for diagnosis or gene therapy, e.g. for treating
CC conditions characterised by hyper-proliferative cells (e.g. tumours),
CC genetic diseases (e.g. Huntington's disease, Tay-Sachs disease, or
CC sickle cell disease), or infections (e.g. HIV infection). They can also
CC be used for in vitro production of biologically active proteins.
SQ Sequence 10610 BP; 2807 A; 2821 C; 2446 G; 2536 T;

Query Match 100.0%; Score 345; DB 1; Length 10610;
Best Local Similarity 100.0%; Pred. No. 8.9e-105;
Matches 345; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atggtcttcacagctcttcctccgtctccgtgtgtgactcgcagaagaatgtgtagt 60
DB 1711 ATGGTCTTCACAGCTCTCCCGCTCCCGTGTGTGACACGAGACGATGTGTAGT 1652
QY 61 tggctgggtgtgcttattctgcggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 120
DB 1651 TGGCTGGGTGTGCTTATTCTGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1392
QY 121 gttacatagaaacccgaagccagggggtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 180
DB 1591 GTTTACATAGAACCCGAGACGAGGGGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGT 1532
QY 181 aactactacacagagagatcttaagcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 240
DB 1531 AACTACTACACAGAGAGATCTTAAGCGGCGAGACGAGACGAGATCTGTGTGTGTGTGTGT 1472
QY 241 cgcacctgt 300
DB 1471 CGCAGCTGT 1412
QY 301 agcacaacacagatctcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 345
DB 1411 AGCAGCAACAGATCTCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1367
RESULT 4
T60559/c
ID 160559 standard; DNA; 32026 BP.
AC 160559;
DE Recombinant adenovirus H5.020Tneo-int.
KM Adenovirus H5.020Tneo-int.; plasmid pKMTVneo-int.; Ad5; Mo-MLV;
KW gene therapy; vector; neomycin resistance; neo gene; ss.
OS Chimeric - Mastadenovirus serotype 5.
PN MO9715679-A1.
PD 01-MAY-1997.
PE 24-OCT-1996; 017176.
PF 27-OCT-1995; US-005942.
PA (UYPE-) UNIV PENNSYLVANIA.
PI Kelley WM, Wilson JM.
PT WPI: 97-259031/23.
DR Recombinant replication defective virus - inserts transgene into
PT host cell chromatin in the presence of transposase, providing stable
PT and durable transgene expression
PS Example 2; Fig 10A-H; 74pp; English.
CC This is the DNA sequence of the recombinant adenovirus
CC H5.020Tneo-int. It was obtained by co-transfecting HEK293 (ATCC
CC CRL1573) cells with linearised plasmid pAdneo-int (see T60555) and
CC Ad5d17001. It was used as a control in studies of the
CC retrotransposition and integration of cis- and trans-acting
CC recombinant adenoviruses (see T60557 and T60558) into host cells.
CC Claimed novel recombinant replication defective viruses (A)
CC comprise: (a) DNA of, or corresponding to, at least part of the
CC viral genome, able to infect a mammalian cell; and (b) a first

CC expression sequence (ES1) comprising a human gene (I) linked to
CC regulatory sequences for expression, with the cassette flanked by
CC the cis-acting terminal repeat sequences of a transposon, and ES1
CC flanked by DNA of (a); (A) can infect a mammalian cell and can
CC express (I) and transfer it to the cellular chromatin in vivo or in
CC vitro in the presence of a transposase. Also new are: (1) a
CC recombinant replication defective virus (B) containing (a) as above
CC and a second expression sequence (ES2) containing a trans-acting
CC transposase gene (II) plus regulatory sequences, flanked by DNA of
CC (a) and able to express transposase in mammalian cells; and (2)
CC mammalian cells which stably express (I) integrated into its
CC chromatin, produced by infection with (A). (A) and (B) are used
CC for delivery and stable integration of (I) into a host cell
CC (claimed), e.g. in somatic gene therapy of genetic defects or
CC deficiencies such as cystic fibrosis. (A) are also used to prepare
CC recombinant retroviruses (RVV) (claimed). Because (I) becomes
CC stably integrated, it provides longer lasting expression than genes
CC introduced with conventional adenoviral vectors and the need for
CC repeated administration is avoided. The transgene is inherited by
CC progeny cells and the viral component is gradually degraded by the
CC cell. When used for RVV production, (A) increase the ratio of RVV
CC to empty retroviruses.
SQ Sequence 32026 BP; 7388 A; 9141 C; 8873 G; 6624 T;

Query Match 100.0%; Score 345; DB 1; Length 32026;
Best Local Similarity 100.0%; Pred. No. 1.4e-104;
Matches 345; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atggtcttcacagctcttcctccgtctccgtgtgtgactcgcagaagaatgtgtagt 60
DB 30433 ATGGTCTTCACAGCTCTCCCGCTCCCGTGTGTGACACGAGACGATGTGTAGT 30374
QY 61 tggctgggtgtgcttattctgcggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 120
DB 30373 TGGCTGGGTGTGCTTATTCTGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 30314
QY 121 gttacatagaaacccgaagccagggggtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 180
DB 30313 GTTTACATAGAACCCGAGACGAGGGGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGT 30254
QY 181 aactactacacagagagatcttaagcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 240
DB 30253 AACTACTACACAGAGAGATCTTAAGCGGCGAGACGAGACGAGATCTGTGTGTGTGTGTGT 30194
QY 241 cgcacctgt 300
DB 30193 CGCAGCTGT 30134
QY 301 agcacaacacagatctcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 345
DB 30133 AGCAGCAACAGATCTCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 30089
RESULT 5
V07261/c
ID V07261 standard; DNA; 34303 BP.
AC V07261;
DE 28-SEP-1998 (first entry)
KM Adenoviral vector plasmid pBHG11.
KW Adenovirus 5; Ad5; vector; gene therapy; ds.
OS Mastadenovirus 5.
FH Key Location/Qualifiers
FT 8772..9385
FT misc_feature
FT /tag- a
FT /note- "region deleted in plasmid pdeprp and
FT /tag- b
FT /note- "region deleted in plasmid pBHG11delprp (Claim 45)"
FT 8773..9586
FT misc_feature
FT /tag- b
FT /note- "region deleted in plasmid
FT pBHG11delprpPRVARN+ct3 and
FT pBHG11delprpPRVARN+ct3 (Claim 53)"
FT 10705..11134
FT misc_feature

Query Match	Best Local Similarity	100.0%	Score 345;	DB 1;	Length 34303;
Matches 345;	Conservative	0;	Mismatches	0;	Indels
0Y	1	atggtcttcagagccttcgcgcctctccgcggtgtgtgactgcagaaagtgttagt	60		
Db	32711	ATGGTCTTCAGAGCCTTCGCCCTCCCGTGTGAGTCCAGAGCAAGTGAAGT	32652		
0Y	61	tggcttggtgtggtctattctgcggtgtgtgagtattatcaaggcagcgcgcatgaaga	120		
Db	32651	TGGCTGGGTGGTGTATTATCTGCGGAGGAGATGATTATCAGGCGCAGCGCATGAAGA	32592		
0Y	121	gtttacatagaaccgcgaagcagggggcgctgtgagtctttagaagagtgtatatactac	180		
Db	32591	GTTTACATAGAACCAGGAGCCGAGGGGCGCTTGATGCTTTACAGAGAGTGATATACATC	32532		
0Y	181	aactactacacagagcgatcttaagcgcgagacccggaagcagaagatcttctgtacgcc	240		
Db	32531	AACCTACTACACAGACGATCTTAAGCGCGCAGACCCGAGACCGAGATCTTGTGTACGCC	32472		
0Y	241	cgcacctggtttgtcttaaganaataatgactacgctcggcggttccatttgcgatgacct	300		
Db	32471	CGCACCTGGTTTGTCTTACAGAAATATGACTACGTCCGGCGTTTCATTGGGATGACACT	32412		
0Y	301	acgaccacaacgattcggtgtgtcttcgagcactccgtacaaatag	345		
Db	32411	ACGACCAACACGATCTCGTGTGTCTTCGCGGACCTCCGTACAACTAG	32367		

ID	Accession	Standard	DNA	34382 BP
AC	X15627	standard	DNA	34382 BP
DT	07-MAY-1999	(first entry)		
DE	Recombinant adenovirus Ad: pac-beta-Gal			
KW	ELa region: E3 region; gene therapy; nitric oxide synthetase; NOS;			
KM	atherosclerotic artery; ss.			
OS	Synthetic.			
PH	Key	Location/Qualifiers		
FT	complement (3372..333)			
FT	CD5			
PD	US5880102-A.			
PF	09-MAR-1999.			
PR	17-JAN-1995; 374483.			
PA	17-JAN-1995; US-374483.			
PI	(UYDU-) UNIV DUKE.			
PI	Blazing MA, George SE;			
DR	WPI: 98-204005/17.			
P	PSDB: W97243.			
CC	New replication deficient adenovirus bearing deletions of the E1a and E3 regions - containing a single packaging signal sequence and E1a enhancer sequence, the E1a deletion has unique cleavage sites and is useful as a gene therapy vector			
PS	Disclosure: Fig 49-100; 148pp; English.			
CC	The present sequence represents recombinant adenovirus			
CC	Ad: pac-beta-gal, which exemplifies the invention. The specification describes replication deficient adenoviruses having deletions of the E1a and E3 regions and comprise a single packaging signal sequence and E1a enhancer sequence, where the sequences are at the 3' end of the adenovirus and the E1a deletion contains at least one PacI, ClaI, XbaI or BstBI cleavage site. The replication deficient viral vectors can be used in gene therapy regimens to effect the transfer of genes encoding molecules of therapeutic importance, including isoforms of the nitric oxide synthetase (NOS) gene (brain, endothelial and microphage NOS), the cystic fibrosis chloride channel (CFTR) gene, the dystrophin gene, the LDL receptor gene and the erythropoietin gene. The NOS isoforms can be used in vascular applications or in cancer therapy (microphage NOS). The NOS gene can be introduced into vein grafts prior to their use as coronary artery bypass grafts. A NOS containing adenovirus can also be used following coronary angioplasty to prevent rethrombosis and to treat atherosclerotic arteries.			
SO	Sequence 34382 BP;	7923 A;	9880 C;	9421 G; 7158 T;
Query Match	100.0%;	Score 345;	DB 1;	Length 34382;
Best Local Similarity	100.0%;	Pred. NO. 1.4e-104;		
Matches 345;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
1	atggtcttccagcgtcttcctccgctccctccggtgtgtgactcgacgaagaatgttagt	60		
2	atggtcttccagcgtcttcctccgctccctccggtgtgtgactcgacgaagaatgttagt	60		
3	atggtcttccagcgtcttcctccgctccctccggtgtgtgactcgacgaagaatgttagt	60		
4	atggtcttccagcgtcttcctccgctccctccggtgtgtgactcgacgaagaatgttagt	60		
5	atggtcttccagcgtcttcctccgctccctccggtgtgtgactcgacgaagaatgttagt	60		
6	atggtcttccagcgtcttcctccgctccctccggtgtgtgactcgacgaagaatgttagt	60		
7	atggtcttccagcgtcttcctccgctccctccggtgtgtgactcgacgaagaatgttagt	60		
8	atggtcttccagcgtcttcctccgctccctccggtgtgtgactcgacgaagaatgttagt	60		
9	atggtcttccagcgtcttcctccgctccctccggtgtgtgactcgacgaagaatgttagt	60		
10	atggtcttccagcgtcttcctccgctccctccggtgtgtgactcgacgaagaatgttagt	60		
11	atggtcttccagcgtcttcctccgctccctccggtgtgtgactcgacgaagaatgttagt	60		
12	atggtcttccagcgtcttcctccgctccctccggtgtgtgactcgacgaagaatgttagt	60		
13	atggtcttccagcgtcttcctccgctccctccggtgtgtgactcgacgaagaatgttagt	60		
14	atggtcttccagcgtcttcctccgctccctccggtgtgtgactcgacgaagaatgttagt	60		
15	atggtcttccagcgtcttcctccgctccctccggtgtgtgactcgacgaagaatgttagt	60		
16	atggtcttccagcgtcttcctccgctccctccggtgtgtgactcgacgaagaatgttagt	60		
17	atggtcttccagcgtcttcctccgctccctccggtgtgtgactcgacgaagaatgttagt	60		
18	atggtcttccagcgtcttcctccgctccctccggtgtgtgactcgacgaagaatgttagt	60		
19	atggtcttccagcgtcttcctccgctccctccggtgtgtgactcgacgaagaatgttagt	60		
20	atggtcttccagcgtcttcctccgctccctccggtgtgtgactcgacgaagaatgttagt	60		
21	atggtcttccagcgtcttcctccgctccctccggtgtgtgactcgacgaagaatgttagt	60		
22	atggtcttccagcgtcttcctccgctccctccggtgtgtgactcgacgaagaatgttagt	60		
23	atggtcttccagcgtcttcctccgctccctccggtgtgtgactcgacgaagaatgttagt	60		
24	atggtcttccagcgtcttcctccgctccctccggtgtgtgactcgacgaagaatgttagt	60		
25	atggtcttccagcgtcttcctccgctccctccggtgtgtgactcgacgaagaatgttagt	60		
26	atggtcttccagcgtcttcctccgctccctccggtgtgtgactcgacgaagaatgttagt	60		
27	atggtcttccagcgtcttcctccgctccctccggtgtgtgactcgacgaagaatgttagt	60		
28	atggtcttccagcgtcttcctccgctccctccggtgtgtgactcgacgaagaatgttagt	60		
29	atggtcttccagcgtcttcctccgctccctccggtgtgtgactcgacgaagaatgttagt	60		
30	atggtcttccagcgtcttcctccgctccctccggtgtgtgactcgacgaagaatgttagt	60		
31	atggtcttccagcgtcttcctccgctccctccggtgtgtgactcgacgaagaatgttagt	60		
32	atggtcttccagcgtcttcctccgctccctccggtgtgtgactcgacgaagaatgttagt	60		
33	atggtcttccagcgtcttcctccgctccctccggtgtgtgactcgacgaagaatgttagt	60		
34	atggtcttccagcgtcttcctccgctccctccggtgtgtgactcgacgaagaatgttagt	60		
35	atggtcttccagcgtcttcctccgctccctccggtgtgtgactcgacgaagaatgttagt	60		
36	atggtcttccagcgtcttcctccgctccctccggtgtgtgactcgacgaagaatgttagt	60		

301 agcagcaacacgacgtcgtgtgtcgtgcgcacccgtacagtag 345
DB 33243 ACGACCAACAGATCTGCTGCTCGGCGACATCCGTACAGTAG 32199

RESULT 7
ID T60557/C
T60557 standard; DNA; 35000 BP.

AC T60557;
DE 05-MAR-1998 (first entry)
KW Adenovirus H5.020tkneo-int(LTR); plasmid pAdMvneo-int;
KW retrovirus; retrotransposition; transposition; transgene;
KW gene therapy; vector; neomycin resistance; neo gene; ss.
OS ChimERIC - Mastadenovirus serotype 5.
OS ChimERIC - Moloney murine leukaemia virus.
OS ChimERIC - Rattus sp.
OS ChimERIC - Synthetic.
PN W0915679-A1.
PD 01-MAY-1997.
PF 24-OCT-1996; U17176.
PR 27-OCT-1995; US-005942.
PA (UYPF-) UNIV PENNSYLVANIA.
PI Kelley MA, Wilson JM.
DR WPI; 97-259031/23.
PT Recombinant replication defective virus - inserts transgene into
PT host cell chromatin in the presence of transposase, providing stable
PT and durable transgene expression.
PS Example 2: Fig 8A-I; 74pp; English.
CC This is the DNA sequence of the cis-acting recombinant adenovirus
CC H5.020tkneo-int(LTR). It was obtained by co-transfecting HEK293
CC (ATCC CRL1573) cells with linearised plasmid pAdMvneo-int
CC (see T60554) and Ad5d17001. A trans-acting adenovirus (see
CC T60558) was also produced, and retrotransposition and integration
CC into host cells was demonstrated in HeLa cells infected with both
CC viruses. Claimed novel recombinant replication defective viruses
CC (A) comprises: (a) DNA of, or corresponding to, at least part of the
CC viral genome, able to infect a mammalian cell; and (b) a first
CC expression sequence (ES1) comprising a human gene (1) linked to
CC the cis-acting terminal repeat sequences, with the cassette flanked by
CC the cis-acting terminal repeat sequences of a transposon, and ES1
CC flanked by DNA of (a). (A) can infect a mammalian cell, and can
CC express (1) and transfer it to the cellular chromatin in vivo or in
CC vitro in the presence of a transposase. Also new are: (1) a
CC recombinant replication defective virus (B) containing (a) as above
CC and a second expression sequence (ES2) containing a trans-acting
CC transposase gene (T1) plus regulatory sequences, flanked by DNA of
CC (a) and able to express transposase in mammalian cells; and (2)
CC mammalian cells which stably express (1) integrated into its
CC chromatin, produced by infection with (A). (A) and (B) are used
CC for delivery and stable integration of (1) into a host cell
CC (claimed), e.g. in somatic gene therapy of genetic defects or
CC deficiencies such as cystic fibrosis. (A) are also used to prepare
CC recombinant retroviruses (RV) (claimed). Because (1) becomes
CC stably integrated, it provides longer lasting expression than genes
CC introduced with conventional adenoviral vectors and the need for
CC repeated administration is avoided. The transgene is inherited by
CC progeny cells and the viral component is gradually degraded by the
CC cell. When used for RRV production, (A) increase the ratio of RRV
CC to empty retrovirus.
SQ Sequence 35000 BP; 7810 A; 9673 C; 9271 G; 7046 T;

Query Match 100.0%; Score 345; DB 1; Length 35000;
Best Local Similarity 100.0%; Pred. No. 1.4e-104;
Matches 345; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atggtcttcacgctctccgcgtccctccggtgtgtactgcagacgaatgtgtagt 60
DB 33407 ATGGTCTTCACGCTCTCCGCGTCTCCGCTGTGTGACTGCGAAGCAATGTGTAGGT 33348
QY 61 tgcgtggtgtgtgcttattctgcggtgtgtgagtattatcagggcagcgcatgaagga 120

DB 33347 TGGCTGGGTGTGCTTATCTCGCGTGTGTGATTTATCAGCGCGCGCATGAAGA 33328
QY 121 gttacatagaacccagagcgagggcgctgtgatatctttgagaggtgatatctac 180
DB 33287 GTTTACATAGAACCCGAAGCCAGGGGCGCTGTGATCTTTGAGAGAGTGTATCTAC 33228
QY 181 aactactacacagagcatctaaagcgagcgagacggagacagcatctgtttgcagcc 240
DB 33227 AACTACTACACAGAGCATCTAAGCGGCGAGACCGGAGACAGCATCTGTTTGCACGCC 33168
QY 241 cgcacctgtgttgcctcaggaatatgactacgctcggcggtccattgtgcatact 300
DB 33167 CGCACCTGCTTGTCTTCAGGAATATGACTACGTCGCGGCTTCATTGGCATGACACT 33108
QY 301 agcagcaacacgacgtcgtgtgtcgtgcgcacccgtacagtag 345
DB 33107 ACGACCAACAGATCTGCTGCTCGGCGACATCCGTACAGTAG 33063

RESULT 8

ID V07258/C
V07258 standard; DNA; 35935 BP.
AC V07258;
DE 28-SEP-1998 (first entry)
KW Adenovirus 5 genome.
KM Adenovirus 5; Ad5; vector; gene therapy; ds.
OS Mastadenovirus 5.
PN W09817783-A1.
PD 30-APR-1998.
PF 23-OCT-1997; U19541.
PR 23-OCT-1996; US-735609.
PA (UNMI) UNIV MICHIGAN.
PI Analifitano A, Chamberlain JS, Hartigan-O'Connor DJ,
PI Husner MA, Kumar-Singhr;
DR WPI; 98-261485/23.
PT New adenoviral recombinant plasmid(s) - comprise sequences provided
PT for expression of large foreign DNA fragments, used for, e.g. gene
PT therapy of genetic disease(s)
PS Disclosure: Page 69-86; 139pp; English.
CC This nucleotide sequence comprises the adenovirus 5 (Ad5) genome.
CC The invention provides improved adenoviral vectors and packaging
CC cell lines. One type of improved vector comprises deletions
CC within the E2b region of the adenoviral genome (see also V07261).
CC These E2b-deleted viruses are used in conjunction with novel cell
CC lines that constitutively express E2b gene products. The invention
CC also provides vectors deleted for all viral coding regions. These
CC 'gutted' vectors permit the transfer of large genes (e.g. up to
CC 35 kb) to cells, as demonstrated by the transfer of the dystrophin
CC gene to the muscle of mice. The E2b-deleted and gutted vectors
CC provide improved adenoviral vectors useful for a wide variety of
CC gene therapy applications.
SQ Sequence 35935 BP; 8367 A; 10074 C; 9760 G; 7734 T;

Query Match 100.0%; Score 345; DB 1; Length 35935;
Best Local Similarity 100.0%; Pred. No. 1.5e-104;
Matches 345; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atggtcttcacgctctccgcgtccctccggtgtgtactgcagacgaatgtgtagt 60
DB 34342 ATGGTCTTCACGCTCTCCGCGTCTCCGCTGTGTGACTGCGAAGCAATGTGTAGGT 34283
QY 61 tgcgtggtgtgtgcttattctgcggtgtgtgagtattatcagggcagcgcatgaagga 120
DB 34282 TGGCTGGGTGTGCTTATCTCGCGTGTGTGATTTATCAGCGCGCGCATGAAGA 34223
QY 121 gttacatagaacccagagcgagggcgctgtgatatctttgagaggtgatatctac 180
DB 34222 GTTTACATAGAACCCGAAGCCAGGGGCGCTGTGATCTTTGAGAGAGTGTATCTAC 34163
QY 181 aactactacacagagcatctaaagcgagcgagacggagacagcatctgtttgcagcc 240

DB 34162 AACCTACTACAGACGAGACTAAGCGGCGAGACCGGAGACGAGACTGTTGTGCAGCC 34103

07 241 cgcacctggttttgcttcaggaataatagctactgctccgagcttccattggatgacct 300
|||||
DB 34102 CGCAGCTGGTTTGTCTCAGGAAATATGATGACTACGTCCGGCGTTCATTGGCATGACACT 34043

07 301 acgacacacacgatcgtgtgtcttcgscgacctccgtacagtag 345
|||||
DB 34042 ACGACCAACAGCATCTCGGTTGTCTCGCGGCGACTCGTACAGTAG 33998

RESULT 9

ID T60558/c
T60558/c
T60558: standard; DNA: 36538 BP.

AC T60558:
DE 05-MAR-1998 (first entry)
KW Adenovirus H5.020CMVag-pol; plasmid pAdCMVneo-Int; Ad5; Mo-MLV;
KW retrovirus; retrotransposition; transduction; transgene;
KW gene therapy; vector; ss.
OS Chimeric - Mastadenovirus serotype 5.
OS Chimeric - Cytomegalovirus.
OS W09715679-Al.
PM 01-MAY-1997.
PM 24-OCT-1996; U17176.
PM 27-OCT-1995; US-005942.
PA (UYPE-) UNIV PENNSYLVANIA.
PI Kelley WM, Wilson JM;
WP1: 97-259031/23.
PT Recombinant replication defective virus - inserts transgene into
PT host cell chromatin in the presence of transposase, providing stable
PS and durable transgene expression.
PS Example 2; Fig 9A-I; 74pp; English.
CC This is the DNA sequence of the trans-acting recombinant adenovirus
CC H5.020CMVag-pol. It was obtained by co-transfecting HEK293
CC (ATCC CRL1573) cells with linearised plasmid pAdCMVag-pol (see
CC T60558) and Ad5d17001. A cis-acting adenovirus (see T60557) was
CC also produced, and retrotransposition and integration into host
CC cells was demonstrated in HeLa cells infected with both viruses.
CC Claimed novel recombinant replication defective viruses (A)
CC comprise: (a) DNA of, or corresponding to, at least part of the
CC viral genome, able to infect a mammalian cell; and (b) a first
CC expression sequence (E51) comprising a human gene (I) linked to
CC regulatory sequences for expression, with the cassette flanked by
CC the cis-acting terminal repeat sequences of a transposon, and E51
CC flanked by DNA of (a); (A) can infect a mammalian cell and can
CC express (I) and transfer it to the cellular chromatin in vivo or in
CC vitro in the presence of a transposase. Also new are: (1) a
CC recombinant replication defective virus (B) containing (a) as above
CC and a second expression sequence (E52) containing a trans-acting
CC transposase gene (II) plus regulatory sequences, flanked by DNA of
CC (a) and able to express transposase in mammalian cells; and (2)
CC mammalian cells which stably express (I) integrated into its
CC chromatin, produced by infection with (A). (A) and (B) are used
CC for delivery and stable integration of (I) into a host cell
CC (claimed), e.g. in somatic gene therapy of genetic defects or
CC deficiencies such as cystic fibrosis. (A) are also used to prepare
CC recombinant retroviruses (RRV) (claimed). Because (I) becomes
CC stably integrated, it provides longer lasting expression than genes
CC introduced with conventional adenoviral vectors and the need for
CC repeated administration is avoided. The transgene is inherited by
CC progeny cells and the viral component is gradually degraded by the
CC cell. When used for RRV production, (A) increase the ratio of RRV
CC to empty retrovirus.
CC Sequence 36538 BP; 8621 A; 10490 C; 9909 G; 7518 T;

Query Match 100.0%; Score 345; DB 1; Length 36538;
Best Local Similarity 100.0%; Pred. No. 1.5e-104;
Matches 345; Conservative 0; Mismatches 0; Indels 0; Gaps 0

1 atggtcttcacgactcttcctccgctccctcccggtgtgactcgagacgaatgtgtagt 60

Db	34945	ATGGTCTTCACACTCTTCCGCTCTCCCGTGTGACTCGACGAGCAAGATGTAGGT	34868
Qy	61	tgcctgggtgtgtgcttatctctcgtgtgtgtatgttatcatcaaggcgcgcgtcatgaaga	120
Db	34885	TGGCTGGGTGTGGCTTATTTCTCGGTGTGTGATGTATACAGGGCAGGGGCGCATGAAGA	34828
Qy	121	gtttacataaacccgaagcgaaggcgccctggaatgtgttgaagaagtgtatactac	180
Db	34825	GTTTACTATTAACCCGACACCGAGGGGCGCCCTCGATGCTTTGAAGAGTGTGATTAACA	34766
Qy	181	aactactacacagacgatctcaagcgcgcgaacgcgagacgcagatctgtttcaagcc	240
Db	34765	AACACTACACACAGACGATCTAAGCGGGGAGACGCGAGACGCACATCTGTTTCTACGCC	34706
Qy	241	cgaaactggtttgtctcagaataatbactacgtccgggttcatttgatgcagcact	300
Db	34705	CGCACTGGGTTTGTCTTACGGAATATGACTACGTCGGCCTTCATTGTCATGACACT	34646
Qy	301	acgacacacacgactcgcgtgttctcggcgacactcgtacagatg	345
Db	34645	ACGACCAACACGATCTCGGTGTGTCTCGGCGCACTCCGATAGTAG	34601
RESULT	10		
ID	X07371/c		
AC	X07371 standard; DNA; 34427 BP.		
DT	07-JUN-1999 (first entry)		
DE	Adenovirus vector 231-10 genome.		
KM	Adenovirus, vector 231-10; apoptosis; inhibitor; RID-alpha;		
KW	RID-beta; receptor internalisation and degradation; cancer;		
KW	degenerative disease; immune disorder; gene therapy;		
KW	tissue transplant; death receptor; tumour necrosis factor receptor;		
KW	human; ds.		
OS	Chimeric - Mastadenovirus 5.		
OS	Chimeric - Mastadenovirus 2.		
OS	Chimeric - Human cytomegalovirus.		
PN	W09902658-AL.		
PD	21-JAN-1999.		
PR	08-JUL-1998; U14239.		
PR	09-JUL-1997; US-088993.		
PA	(UYSL-) UNIV SAINT LOUIS.		
PI	MOLD MSW.		
DR	WPI: 99-120862/10.		
PT	Method for inhibiting apoptosis - using receptor internalisation and		
PT	degradation (RID) complex		
PS	Example 10; Fig 28A-XX; 126PP; English.		
CC	This is the nucleotide sequence of vector 231-10, a human		
CC	adenovirus serotype 5 (Ad5) vector. 231-10 lacks the E1A, E1B		
CC	and E3 transcription units. The deleted E1A and E1B regions are		
CC	replaced with an expression cassette wherein all the E3 proteins		
CC	are expressed from the human cytomegalovirus (CMV) promoter. The		
CC	E3 expression cassette contains the E3 genes from virus pm734.1,		
CC	a derivative of virus rec700, which is an Ad5-Ad2-Ad recombinant		
CC	that has the Ad2 version of the E3 genes for the 12.5k, 6.7k,		
CC	gpl9k and RID-alpha proteins, and the Ad5 version of the E3 genes		
CC	for the RID-beta and 14.7k proteins. In addition, the vector has		
CC	2 missense mutations in the adp (adenovirus death protein) gene		
CC	that eliminate the first 2 Met codons, thereby precluding synthesis		
CC	of functional ADP. Because 231-10 lacks E1A, viral genes in the		
CC	vector backbone are not expressed; only the E3 proteins are		
CC	expressed from the CMV promoter. Thus, the vector serves as an		
CC	essentially inert vehicle by which E3 genes can be delivered into		
CC	a cell. The invention provides a method for inhibiting apoptosis of		
CC	a cell expressing a death receptor of the tumour necrosis factor		
CC	receptor family. This involves treating the cells with a receptor		
CC	internalisation and degradation (RID) protein complex containing		
CC	RID-alpha (10.4 k, see W97877 and W97878) and RID-beta (14.5k, see		
CC	W97879) proteins encoded by the E3 region of adenovirus. The cell		
CC	can be treated by administering to the cell a polynucleotide		
CC	expressing the RID complex or a composition containing the RID		
CC	complex. The compositions and methods are used to treat		

CC degenerative immunodeficiency diseases, particularly to decrease
 CC leukocyte apoptosis (claimed). Immune disorders that can be
 CC treated include autoimmune disorders, ischemic neuron death as
 CC caused by myocardial infarction, stroke induced neuron death and
 CC reperfusion injury, alcohol-induced hepatitis, diseases caused by
 CC viral infection such as AIDS and fulminant hepatitis, and cancer.
 CC The methods can also be used to promote tissue transplant
 CC survival. Vector 231-10 can be used to deliver the RID complex to
 CC the cells.
 SQ Sequence 34427 BP; 8098 A; 9606 C; 9451 G; 7272 T;

Query Match 99.5%; Score 343.4; DB 1; Length 34427;
 Best Local Similarity 99.7%; Pred. No. 4.9e-104;
 Matches 344; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 atggtcttcacagctctccgcgtccctccggtgtgtgactcgacgaacgaatgtgtagt 60
 |||||
 DB 32834 ATGGTCTTCCAGCTCTCCCGCTCCCTCCGTCGTGACTCGACAGAAATGCTAGCT 32775
 |||||
 QY 61 tggctgggtgtggtctatctcgcgtgtgtgtgtatctacagggcgagcgagcaagga 120
 |||||
 DB 32774 TGGCTGGGTGTGGCTTATTCGCGTGTGATGTTATCAGGGCCGCGCATGAGGA 32715
 |||||
 QY 121 gttacatagaacccgaagcagagggcgctgtgactgtcttgaagagtgatatactac 180
 |||||
 DB 32714 GTTATACATAGAACCCGAGCCAGGGGGGCGCTGATGCTTGAGAGAGTGAATATAC 32655
 |||||
 QY 181 aactactacacagagcagatctaaagcgagagcagagcagagcagatctgttgcagcc 240
 |||||
 DB 32654 AACTACTACACAGAGCAGATCTAAAGCGGAGACCGAGACGAGATCTGTTGTACGCC 32595
 |||||
 QY 241 cgcacctgtgttcttaagaaatactactacgtccgscgttccattgtgcatgacct 300
 |||||
 DB 32554 CGCAGCTGTGTTCTTATGAGAAATATGACTACGTCCGCGCTTCATTGGCATACACT 32555
 |||||
 QY 301 acgaccacacagatctcgttctcgcgcagcctccgtacagtag 345
 |||||
 DB 32534 ACGACCAACAGATCTCGGTTCTCGGCGCAGCTCCGTACAGTAG 32490
 |||||

RESULT 11
 V22141/c
 ID V22141 standard; cDNA; 36519 BP.
 AC V22141
 DT 17-AUG-1998 (first entry)
 DE Chimpanzee adenovirus C68 genomic sequence.
 KW genomic sequence; viral genome; virus; gene therapy; treatment;
 KW acquired defects; inherited defects; genetic engineering; vector;
 KW in vitro production; recombinant protein; ds.
 OS Chimpanzee adenovirus C68.
 FT Key Location/Qualifiers
 FT CDS 480..1521
 FT /tag- a
 FT /note- "E1A gene"
 FT 1560..3956
 FT /tag- b
 FT /note- "E1B gene"
 FT complement(23370..21787)
 FT /tag- c
 FT /note- "E2A gene"
 FT complement(10346..3957)
 FT /tag- d
 FT /note- "E2B gene"
 FT 26806..31877
 FT /tag- e
 FT /note- "E3 gene"
 FT complement(36193..33486)
 FT /tag- f
 FT /note- "E4 gene"
 FT 10823..13817
 FT /tag- g
 FT /note- "L1 gene"

FT CDS 13884..17431
 FT /tag- h
 FT /note- "L2 gene"
 FT 17480..21804
 FT /tag- i
 FT /note- "L3 gene"
 FT 23399..27439
 FT /tag- j
 FT /note- "L4 gene"
 FT 32134..33502
 FT /tag- k
 FT /note- "L5 gene"
 FT PN WO9810087-A1.
 FT PD 12-MAR-1998.
 FT PF 04-SEP-1997; U15694.
 FT PR 06-SEP-1996; US-024700.
 FT PA (UYPE-) UNIV PENNSYLVANIA.
 FT PI Farina SF, Fisher KJ, Wilson JM;
 FT WPI: 98-193635/17.
 PT Chimpanzee adenovirus vector - useful in gene therapy and genetic
 PT engineering in general
 PS Disclosure: Pages 70-89; 116pp; English.
 CC The sequence is that of the Chimpanzee adenovirus C68. It may be
 CC used in the production of a vector comprising it and a selected
 CC heterologous gene operatively linked to regulatory sequences
 CC directing its expression in a heterologous cell.
 CC Such recombinant adenoviruses are useful as vectors in gene
 CC therapy and genetic engineering in general, especially for
 CC the treatment of acquired or genetically inherited defects.
 CC The viruses are also useful for the in vitro production of
 CC recombinant proteins of interest.
 SQ Sequence 36519 BP; 7809 A; 10790 C; 10623 G; 7297 T;

Query Match 41.7%; Score 143.8; DB 1; Length 36519;
 Best Local Similarity 64.0%; Pred. No. 1e-37;
 Matches 217; Conservative 0; Mismatches 122; Indels 0; Gaps 0;
 QY 1 atggtcttcacagctctccgcgtccctccggtgtgtgactcgacgaacgaatgtgtagt 60
 |||||
 DB 34942 ATGGTCTTCCGTCTTCTCCCTCCCGCGCTCAACGAAACCAAAATGCTATTC 34883
 |||||
 QY 61 tggctgggtgtggtctatctcgcgtgtgtgtgtatctacagggcgagcgagcaagga 120
 |||||
 DB 34882 TGGCTGGGTGTGGCTTATTCGCGTGTGATGTTATCAGGGCCGCGCATGAGGA 34823
 |||||
 QY 121 gttacatagaacccgaagcagagggcgctgtgactgtcttgaagagtgatatactac 180
 |||||
 DB 34822 ATCTTCACTACCTCCGAGGCGCTTAGATGTGTCAGCGTCTGAGAGACTGCTTTAC 34763
 |||||
 QY 181 aactactacacagcagatctaaagcgagagcagagcagagcagatctgttgcagcc 240
 |||||
 DB 34762 AATTCAACACGAGCGATTCACAGCGCGAGACCGCCGCGATCGGTGTGACGCC 34703
 |||||
 QY 241 cgcacctgtgttcttaagaaatactactacgtccgscgttccattgtgcatgacct 300
 |||||
 DB 34702 CGGACCAAGTCTGCTTACAGAAATATGTCGAAACCAATCATCATGACACG 34643
 |||||
 QY 301 acgaccacacagatctcgttctcgcgcagcctccgtacagtag 339
 |||||
 DB 34642 GTCCGGAACAGATCAGCCGCTCCCGCATCACCCGTA 34604
 |||||

RESULT 12
 V22140/c
 ID V22140 standard; cDNA; 35524 BP.
 AC V22140
 DT 17-AUG-1998 (first entry)
 DE Chimpanzee adenovirus C1 genomic sequence.
 KW genomic sequence; viral genome; virus; gene therapy; treatment;
 KW acquired defects; inherited defects; genetic engineering; vector;
 KW in vitro production; recombinant protein; ds.
 OS Chimpanzee adenovirus C1.

Key	Location/Qualifiers
Key	480..1540
CDS	
FT	/tag- a
FT	/note- "E1A gene"
FT	1566..3958
FT	/tag- b
FT	/note- "E1B gene"
FT	complement(23665..22065)
FT	/tag- c
FT	/note- "E2A gene"
FT	complement(10379..3959)
FT	/tag- d
FT	/note- "E2B gene"
FT	27181..31375
FT	/tag- e
FT	/note- "E3 gene"
FT	complement(35228..32535)
FT	/tag- f
FT	/note- "E4 gene"
FT	10893..13864
FT	/tag- g
FT	/note- "L1 gene"
FT	13925..17591
FT	/tag- h
FT	/note- "L2 gene"
FT	17641..22083
FT	/tag- i
FT	/note- "L3 gene"
FT	23697..27813
FT	/tag- j
FT	/note- "L4 gene"
FT	31556..32551
FT	/tag- k
FT	/note- "L5 gene"
PN	WO9810087-A1.
PD	12-MAR-1998.
PF	04-SEP-1997: U15694.
PR	06-SEP-1996: US-024700.
PA	(TYPE-) UNIV PENNSYLVANIA.
PI	Farina SF, Fisher KJ, Wilson JM;
PI	WPI: 98-193635/17.
PT	Chimpanzee adenovirus vector - useful in gene therapy and genetic
PT	engineering in general
PS	Disclosure: Pages 51-70; 116pp; English.
CC	The sequence is that of the chimpanzee adenovirus C1. It may be
CC	used in the production of a vector comprising it and a selected
CC	heterologous gene operatively linked to regulatory sequences
CC	directing its expression in a heterologous cell.
CC	Such recombinant adenoviruses are useful as vectors in gene
CC	therapy and genetic engineering in general, especially for
CC	the treatment of acquired or genetically inherited defects.
CC	The viruses are also useful for the in vitro production of
CC	recombinant proteins of interest.
CC	Sequence 35524 BP; 8873 A; 9142 C; 9033 G; 8476 T;
Query Match	29.9%; Score 103; DB 1; Length 35524;
Best Local Similarity	60.6%; Pred. NO. 3.6e-24;
Matches 169; Conservative	0; Mismatches 110; Indels 0; Gaps 0;
OY	1 atggtcttcacagccttcctccgtctccctccgtgtgtgactcgacagcaaatgtatagt 60
Db	33994 ATGGTCTTCCATGCTCTCCCTCCCTCTTGATTTAAACCTCAAGCTAATTGATGCA 33935
OY	61 tggctgggtgtgtgctattctcggtgtgtgtatcatcagggcagcgcgcatgaagga 120
Db	33934 TGGTGGGATGATGCCCATTTCTACAGTGAATGATTTTATCAGAGCGATCAGACATGATGGA 33875
OY	121 gttacatagaagaccggaagcgagggggcgctcgatgtctttagagagtgatatactac 180
Db	33874 GTTTTCATACAACTGAGAGCGAGGAACGTTTGACCAACTCAGGGGAATGCGTGTACTTC 33815
OY	181 aactctcacagagcgatctcaagcgcgagacgagcgagagcagagatctgtttgttaagcc 240

Dn	33814	AATTACCACACTAATGCGTTAGGCACGGCATAGAAAGCCGAAGAAGTGTGTCT	33755
Oy	241	cgcactggttttcctcgaatatgctacgtccg	279
Db	33754	AMGCTAGGTTTGTCATGCCAAAATACGAAGATGTCAGG	33716
 RESULT 13 068003/c			
ID	068003	standard; DNA; 36335 BP.	
AC	068003;		
DT	26-MAR-1996	(first entry)	
DE	Ad2/-ORF6/-RGK-CFR nucleotide sequence.		
KW	Recombinant adenovirus; Ad2/CFR-1; adenovirus 2 serotype; E1a; E1b;		
RW	viral replication; gene expression; gene therapy; cystic fibrosis;		
KW	cystic fibrosis transmembrane conductance regulator; CFRP;		
KM	promoter; E3; p15; MHC; class I; viral latency; pulmonary airway; ds.		
OS	Synthetic.		
FT	Key	Location/Qualifiers	
FT	misc_feature	12915..36335	
FT	/tag= a	/note= "Represents residues 10676-34096 of Ad2-E4/ORF6"	
FT	35069..35973		
FT	/tag= b	/note= "Represents residues 33178-34082 of Ad2"	
FT	12915..35054		
FT	/tag= c	/note= "Represents residues 1-32815of Ad2"	
FT	28478..28790		
FT	/tag= d	/number= 1	
FT	/note= "33k protein"	28791..28992	
FT	/tag= e	/number= 1	
FT	28993..29366		
FT	/tag= f	/number= 2	
FT	/note= "33k protein"	13279..14526	
FT	/tag= g	/product= 52,55k protein	
FT	14547..16304		
FT	/tag= h	/product= IIIR protein	
FT	/note= "Peripentonal hexon-associated protein"	16331..16336	
FT	/tag= i	/note= "Major late mRNA L1 poly A signal (putative)"	
FT	16390..18105		
FT	/tag= j	/product= Penton protein	
FT	/note= "Viron component III"	18112..18708	
FT	/tag= k	/product= Pro-VII protein	
FT	/note= "Precursor to major core protein"	18778..19887	
FT	/tag= l	/product= PV protein	
FT	/note= "minor core protein"	20188..20193	
FT	/tag= m	/note= "major late mRNA L2 poly-A signal (putative)"	
FT	20240..20992		
FT	/tag= n	/product= PVI protein	
FT	/note= "Hexon associated precursor"	21077..23983	
FT	/tag= o	/product= Hexon protein	
FT	/note= "Viron component II"		

polya_signal 24657..24662
 /tag- p
 /note- "Major late mRNA L3 poly-A signal (putative)"
 cds complement (244729..26318)
 /tag- q
 /product- DBP protein
 /note- "DNA binding or 72k protein"
 26347..28764
 /tag- i
 /product- 100K protein
 /note- "Hexon assembled"
 29454..30137
 /tag- s
 /product- pVIII protein
 /note- "Hexon-associated precursor"
 30444..30449
 /tag- t
 /note- "Major late mRNA L4 poly-A signal (putative)"
 31051..31530
 /tag- u
 /product- E3 19k protein
 /note- "Glycosylated membrane protein"
 31707..32012
 /tag- v
 /product- E3 11.6k protein
 32008..32013
 /tag- w
 /note- "E3-1 mRNA poly-A signal (putative)"
 33081..33086
 /tag- x
 /note- "E3-2 mRNA poly-A signal (putative)"
 35013..35018
 /tag- y
 /note- "Major late mRNA L5 poly-A signal (putative)"
 1..12914
 /tag- z
 /note- "Represents residues 1-12914 of pAd2/Pgk-CFTR"
 380..914
 /tag- aa
 /note- "pgk promoter"
 1011..5453
 /tag- ab
 /product- CFTR

cds
 W09412649-A.
 09-JUN-1994.
 02-DEC-1993; U11667.
 03-DEC-1992; US-985478.
 01-OCT-1993; US-130682.
 13-OCT-1993; US-136742.
 (GENE) GENZYME CORP.
 PI Armentano D, Couture LA, Gregory RJ, Smith AE;
 DR WPI; 94-200277/24.
 PT Adeno:Virus-based gene therapy vectors - esp. useful for gene
 therapy of cystic fibrosis
 PS Example 15; Page 84-95; 167p; English.
 CC This sequence represents the nucleotide sequence of the second generation
 CC adenovirus vector named Ad2-ORF6/Pgk-CFTR. This virus lacks E1 and in its
 CC place contains a modified transcription unit with the phosphoglycerate
 CC kinase (PGK) promoter and a poly A addition site flanking the cystic
 CC fibrosis transmembrane conductance regulator (CFTR) cDNA. The Pgk
 CC promoter is only if moderate strength but it is long lasting and is not
 CC subject to shut off. The E4 region of the vector has also been modified
 CC in that the whole sequence has been removed and replaced by ORF6, the
 CC only E4 gene essential for growth of adenovirus (Ad) in tissue culture.
 CC The DNA construct comprises a full length copy of the Ad2 genome from
 CC which the early region 1 genes (E1 genes) have been deleted and replaced
 CC by an expression cassette encoding CFTR. The expression cassette
 CC includes the promoter for Pgk and a poly-A addition signal from the
 CC bovine growth hormone gene. The Ad2-ORF6/Pgk-CFTR construct differs
 CC from that given in Q68002 (Ad2/CFTR-1), in that the latter utilises the
 CC endogenous E1a promoter, has no poly-A addition signal directly
 CC downstream of CFTR and retains an intact E4 region. This adenovirus may
 CC be administered to the pulmonary airways in the gene therapy of cystic

CC fibrosis.
 SQ Sequence 36335 BP; 8597 A; 10000 C; 9786 G; 7952 T;
 Query Match 24.8%; Score 85.4; DB 1; Length 36335;
 Best Local Similarity 98.9%; Pred. No. 2.5e-18;
 Matches 86; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 259 aggaataatgactacgtccgcgttcacattgagcatgacactgacacacgactctcg 318
 DB 35975 ATGAATATGACTACGTCCGGCGTTCATTTGGCATGACACTGACACACGACTCTCG 35916
 QY 319 gtgtctcggtcgacactcgtacagtag 345
 DB 35915 GTGTCTCGGCGCACCTCCGTACAGTAG 35889

RESULT 14
 ID T59270
 AC T59270 standard; CDNA; 3653 BP.
 DT 27-AUG-1997 (first entry)
 DE Plasmid pMUTVE4ORF6 minigene sequence.
 KW Minigene; human adenovirus type 5; open reading frame; promoter; in vivo;
 KW mouse mammary tumour virus; growth hormone; terminator; recombinant;
 KW packaging cell line; adeno-associated virus; ex vivo; gene therapy;
 KW inherited disease; cancer; genetic dysfunction; cystic fibrosis; ds.
 OS Synthetic.
 FH Key
 FT Location/Qualifiers
 FT 1..1506
 FT promoter
 FT /tag- a
 FT /note- "mouse mammary tumour virus promoter"
 FT 1521..2405
 FT /tag- b
 FT /product- human Ad5 E4 region ORF6 protein
 FT /note- "specification gives this CDS at nucleotides
 FT 1523-2408"
 FT /transl_except- (pos: 1863..1865, aa: Tyr)
 FT /transl_except- (pos: 2091..2093, aa: Tyr)
 FT 2409..3654
 FT /tag- c
 FT /note- "growth hormone gene terminator sequence"

terminator
 W09639530-A2.
 12-DEC-1996.
 04-JUN-1996; U10245.
 05-JUN-1995; US-462014.
 27-OCT-1995; US-549489.
 (UYPE-) UNIV PENNSYLVANIA.
 PI Fisher KT, Gao G, Wilson JM;
 DR WPI; 97-043152/04.
 DR P-PSDB; W011729.
 PT Recombinant adeno-associated virus contg. second gene which
 PT facilitates its conversion from single stranded to double stranded
 PT virus - enhances efficiency of ex vivo transduction into target cell
 PS Example 1; Page 69-72; 131p; English.
 CC This is the nucleotide sequence of a minigene containing the human
 CC adenovirus type 5 (Ad5) E4 region open reading frame (ORF) 6 under
 CC control of the mouse mammary tumour virus promoter and a growth hormone
 CC terminator sequence. The minigene is used to generate a packaging cell
 CC line expressing the E4 ORF6 product which is used to rescue and package
 CC an E4 deficient recombinant adeno-associated virus (rAAV). The rAAV
 CC comprises at least 2 selected genes, and the first gene able to express the
 CC gene of interest in the target cell; and the second gene able to convert
 CC the single stranded virus to its double stranded form upon expression.
 CC The rAAV is useful in pharmaceutical compositions for use in ex vivo and
 CC in vivo gene therapy for the treatment of inherited diseases, cancer and
 CC other genetic dysfunctions, e.g. cystic fibrosis.
 SQ Sequence 3653 BP; 860 A; 851 C; 933 G; 963 T;

Query Match 23.2%; Score 80; DB 1; Length 3653;
 Best Local Similarity 100.0%; Pred. No. 6.2e-17;
 Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 266 atgactacgtccggcgttcattggcagctacgacacacacacgactcgtttct 325
DB 1521 ATGACTACGTCGGCGGCTTCATTGGCATGACACTACGACCAACAGACTCGTTGCT 1580
QY 326 cggcgacactccgtaacagtag 345
DB 1581 CGCGCACTCCGCTACAGTAG 1600

RESULT 15

TS9273
ID TS9273 standard; DNA; 8299 BP.

AC TS9273; (first entry)

DE 27-AUG-1997 Second generation recombinant adenoviral vector pAV.CMVLP.GRE-ORF6.

RT mouse mammary tumour virus; growth hormone; terminator; recombinant;

RT packaging cell line; adeno-associated virus; ex vivo; gene therapy;

RT inherited disease; cancer; genetic dysfunction; cystic fibrosis; ds.

OS Synthetic.

FE Key

FE repeat_unit

FE repeat_unit

FE repeat_unit

FE repeat_unit

FE repeat_unit

FE repeat_unit

FE repeat_unit

FE repeat_unit

FE repeat_unit

FE repeat_unit

FE repeat_unit

FE repeat_unit

FE repeat_unit

FE repeat_unit

FE repeat_unit

FE repeat_unit

FE repeat_unit

FE repeat_unit

FE repeat_unit

FE repeat_unit

FE repeat_unit

FE repeat_unit

FE repeat_unit

FE repeat_unit

FE repeat_unit

FE repeat_unit

FE repeat_unit

FE repeat_unit

FE repeat_unit

FE repeat_unit

FE repeat_unit

FE repeat_unit

FE repeat_unit

FE repeat_unit

FE repeat_unit

FE repeat_unit

FE repeat_unit

FE repeat_unit

FE repeat_unit

FE repeat_unit

FE repeat_unit

FE repeat_unit

FE repeat_unit

FE repeat_unit

FE repeat_unit

FE repeat_unit

FE repeat_unit

CC virus.
SQ Sequence 8299 BP; 1816 A; 2130 C; 2108 G; 1963 T;
Query Match 23.2%; Score 80; DB 1; Length 8299;
Best Local Similarity 100.0%; Pred. No. 8.7e-17;
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 266 atgactacgtccggcgttcattggcagctacgacacacacacgactcgtttct 325
DB 3402 ATGACTACGTCGGCGGCTTCATTGGCATGACACTACGACCAACAGACTCGTTGCT 3461
QY 326 cggcgacactccgtaacagtag 345
DB 3462 CGCGCACTCCGCTACAGTAG 3481

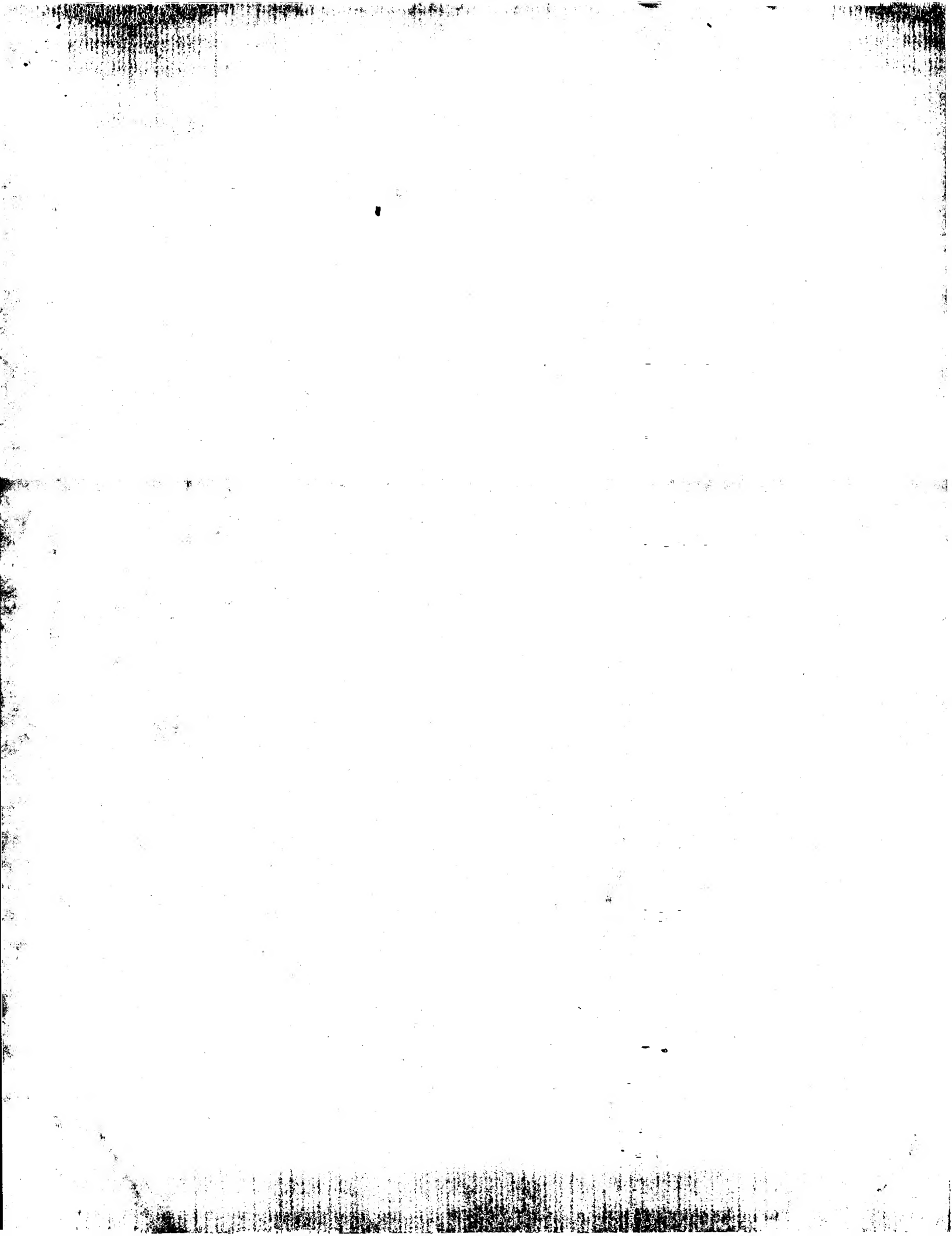
Search completed: June 23, 2000, 10:06:58
Job time: 3686 sec

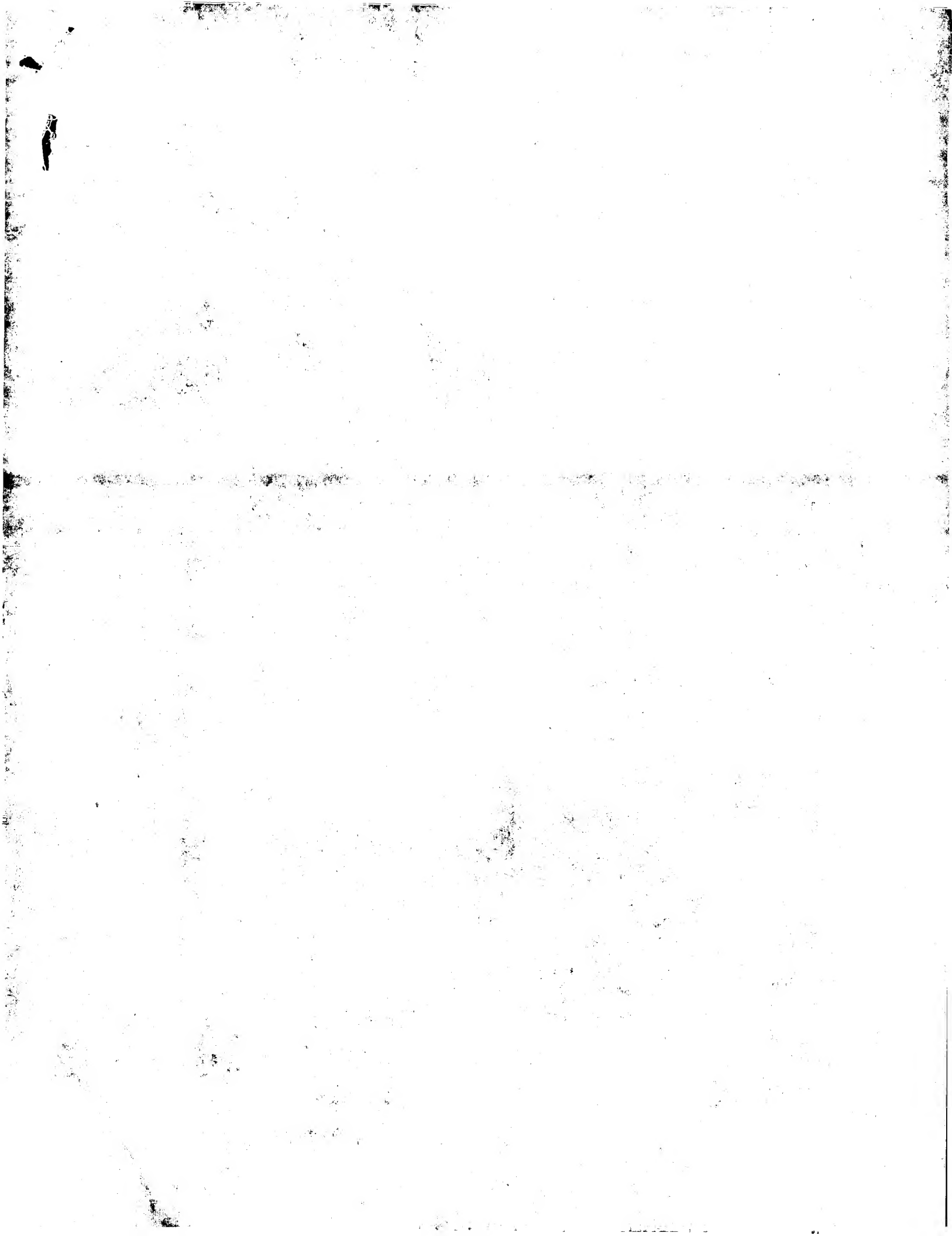
FT repeat_unit 53.219
FT repeat_unit /tag= a
FT repeat_unit /rpt_type= inverted
FT repeat_unit /note= "5' AAV ITR"
FT promoter 255.848
FT promoter /tag= b
FT promoter /function= enhancer promoter
FT promoter /note= "cytomegalovirus enhancer/promoter"
FT cds 914.2892
FT cds /tag= c
FT cds /product= human placental alkaline phosphatase
FT poly_a_signal 2893.3090
FT poly_a_signal /tag= d
FT promoter 3114.3393
FT promoter /note= "SV40 poly-A sequence"
FT promoter /tag= e
FT cds 3402.4286
FT cds /note= "glucocorticoid dependent promoter GRE"
FT poly_a_signal 4315.4713
FT poly_a_signal /tag= f
FT repeat_unit 4547.4713
FT repeat_unit /note= "SV40 poly-A sequence"
FT repeat_unit /tag= g
FT repeat_unit /rpt_type= inverted
FT repeat_unit /note= "3' AAV ITR"
FT misc_feature 4714.8299
FT misc_feature /tag= h
FT misc_feature /note= "plasmid derived sequences"
PD MO9639330-A2.
PD 12-DEC-1996
PD 04-JUN-1996 U10245
PD 05-JUN-1995 US-462014.
PD 27-OCT-1995 US-549489.
PD (TYPE=) UNIT PENNSYLVANIA.
PD Fisher KJ, Gao G, Wilson JM;
PD WPI: 97-043152/04.
PD Recombinant adeno-associated virus contg. second gene which
PD facilitates its conversion from single stranded to double stranded
PD virus - enhances efficiency of ex vivo transduction into target cell
PD Example 19; Page 108-114; 131pp; English.
CC This is the nucleotide sequence of a second generation of recombinant
CC adeno-associated virus (rAAV) which contains 2 genes, the first gene
CC able to express the gene of interest (beta-galactosidase) in the target
CC cell, and the second gene(Ad5 E4 ORF6) able to convert the single
CC stranded virus to its double stranded form upon expression. The rAAV is
CC useful in pharmaceutical compositions for use in ex vivo and in vivo gene
CC therapy for the treatment of inherited diseases, cancer and other genetic
CC dysfunctions, e.g. cystic fibrosis.
CC See also T59271 for another example of a recombinant adeno-associated

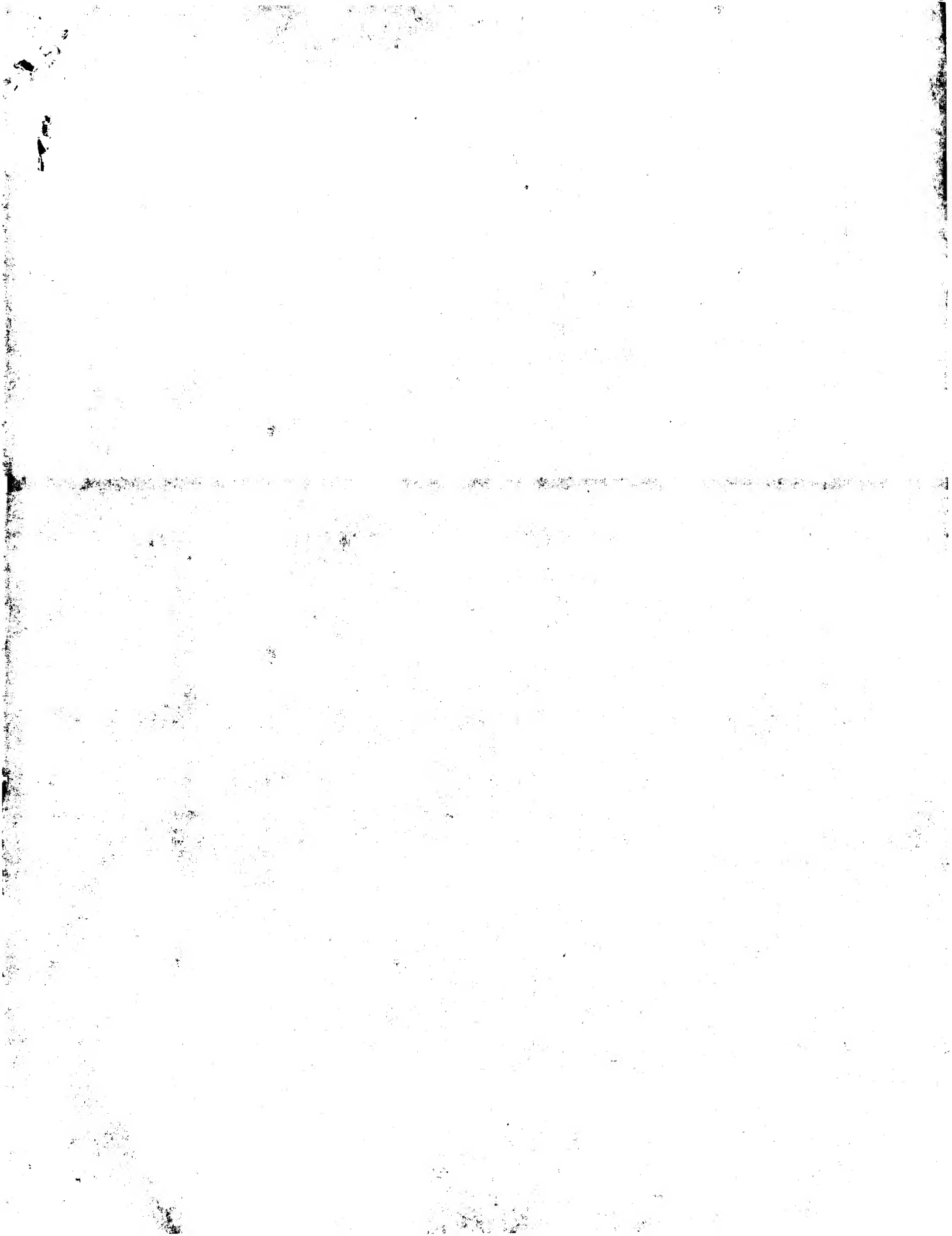
Mon Jun 26 09:10:45 2000

us-09-214-478-3.rng

Page 11







```

source
1..35935
/organism="Human adenovirus type 5"
/db_xref="taxon:28285"
<11565..12297
/note="52/55 k protein gene"
/codon_start=2
/protein_id="AA96406.1"
/db_xref="GI:209843"
/translation="SENHHVRTLVAEEVATIGIMHIMDEVSALEPONPSKPLMAQLF
IGVSHNIEFRDALNIYEPGRWLLDINLOSIVQERSLADYVAIINISML
LGFYVARKIYHPYPIVDKEVKTEGYSYMAKALVITLSDDLYVNNRIHKAVSYSRR
RELSDSEIHSJLORALAGTSGSDREAESEITFDAGADLRWAPSRALAAAGPGLAVAA
AARGNIGCVBEIDEDDEPEDEGEY"
12318..14075
/codon_start=1
/product="protein iIia"
/protein_id="AA96407.1"
/db_xref="GI:289094"
/translation="MMQADTPVRAALOSQSGINSDPROMQDRIMSTLTANPDN
FRQOPANLSALILEAVYDAPARAPPHETAYLAIVNALENRATIRDEAGLYTDALLQEV
AARNSVQNTDRLKCYGVREVAORERKQOQNGISVANAALSTQPAVNPQGDSD
YINPVSALAMTETIPUSEVYISGDPYFQFSRQGLQVNLISQAFKNIQGLMGVRAPT
GDAAVYSLITNSRLTLILLAPFTDSSVSVDYITGLITLTYREALIGQAHVDHTQ

```

PGFGTGGFEVPEGNDGFLMDDIDDSVSPOTLLELQOREDAALRESFRASS
 LSDIGAAPRSDASSPESLSISLTSRTTRPDLGEEYLNNLSLQPOREKNLPAP
 PNNGISLVDKGRMKTYAOEHRDVPDRPPTRRQHHRQRLGWEDDDSDADSSVLD
 TGGSESLVNRDPTQDND

CDS	16286. .16474
-----	---------------

```

/product="protein pVII"

```

```
/protein_id=AA096408.1"
/db_xref="GI:209845"
```

/translation="MRARRLAGIVTPPRSRRRAAAAAAISAMTQGRGNVYW
PDSVSCI PYDVBBDNN"

CDS	16544.	.17650
-----	--------	--------

```

/ product="protein v"

```

```
/procen_id= AAA98409.1"
/db_xref="GI:209846"
```

/translation-MSKRRRIKEEMLOVIAPEIYGPCKEEQDYKPKRLKRVKKKKKDD
DELDDEVELHATABPBYOGKCDYRWRDGGTITTTGAGGAGGAGGAGGAGGAGGAGG

DEDLLEQANERLGEFAYGKRHKMDLALPLDEGNPTPSLKPVTLQOVLPALAPSEKRG

QIPPTSSSIATATEGMETQTSVPASAVADA VQAVAAASKTSTEVOTDPMFRVSA

RQPVLAISVRVAREGGRTLVLPYABHPSTV" ?

```
11114: 11920
/codon start=1
```

```
/product="protein 5 precursor"
```

/db_xref="GI:209847"

[illegible]

```
18003. 18/53
/codon start=1
```

```
/product="protein pVI"  
/protein 14="M100001111"
```

```

/db_xref="GI:209848"

```

KNFGSTVKNYGSKAWNSSTGOMLRDKLKEONFOOKVYDGLASGTSGYVDI.ANNAVONT

ANSALDFKPEVEPPPAVEIVSPEGKGEKRPREDREETLVTVQIDEPSPSYEALKQGLP
TTRPIAPMATGVLGOHTPVTLIDLPAPADTOKPVI.PGBTAYVYTBPSBSGCTNNVACCS

KSLRPVASGNWQSTLNSIVGLGVQSLKRRRCF*
 <25819 26484
 CDS

```
/note="100 K protein"  
/nodes start_i
```

```
/protein_id="AAA96412.1"
```

```

/uu_arel= 01.205049
/translation="ESOTICTELQCPSPDEKSAABCIKITBCIWHCAVIBREHNEU"

```

HAHELKFEYEDSRPPNAELTACVITQGHILGOLAINKARQEFLLRKGCVLLDPQSG
FEINPTPPBPBOYOOBPATASQDQTEF.....TNRD.....

“COURT REPORTERS: THE FUTURE IS NOW”

BASE COUNT	8367 a	10073 c	9761 g	7734 t
ORIGIN				

DB 36342 ATGGTTCCTCCAGCTCTTCCCGCTCCTCCCGTGTGTGACTCGCAGACGATGTAGGT 34283

61. **tgctctggtgctggtctattctcgcggtggtgtaatgtaacagagcgcgcgcatgaagga** 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526
 527
 528
 529
 530
 531
 532
 533
 534
 535
 536
 537
 538
 539
 540
 541
 542
 543
 544
 545
 546
 547
 548
 549
 550
 551
 552
 553
 554
 555
 556
 557
 558
 559
 560
 561
 562
 563
 564
 565
 566
 567
 568
 569
 570
 571
 572
 573
 574
 575
 576
 577
 578
 579
 580
 581
 582
 583
 584
 585
 586
 587
 588
 589
 590
 591
 592
 593
 594
 595
 596
 597
 598
 599
 600
 601
 602
 603
 604
 605
 606
 607
 608
 609
 610
 611
 612
 613
 614
 615
 616
 617
 618
 619
 620
 621
 622
 623
 624
 625
 62

Ob 3422 GTTTACTAGAACCCAGACCAAGGGGGCCCTGGATGCTTGAGAGATGTATTACTAC 3416

Db 34162 AACCTACTACACAGGAGCTTAAGGGCCAGACGAGACACAGATCTGTTTCACGCC 3410:

D03689.1
DB 34102 CGCACCGGTTTGGCTTCAGAGAAATTGACTACGTCCGGCGTTCATTTGGCATTGACATGACACT 3404

Oy 301 acgacacacacgactcgcgttctctcgcgcacctccgacacgtag 345
 |||||
 Db 34042 ACgACCAACGATCTCGGTGTCTCGGGCGACCTCCGTACAGTAg 33998
 .

RESULT 3
ADRCG/c

LOCUS	ADRCG	35937	DP	DNA	VRL	LA-MAN-1730
DEFINITION	Adenovirus type 2, complete genome.					
ACCESSION	J01917	J01918	J01919	J01920	J01921	J01922
	J01923	J01924	J01925	J01926	J01927	J01928
	J01929	J01930	J01931	J01932	J01933	J01934

J01995 J01336 J01937 J01938 J01939 J01940 J01941 J01942 J01943
 J01944 J01945 J01946 J01947 J01948 J01949 J01950 J01951 J01952
 J01953 J01954 J01955 J01956 J00086 K00394 K00395 K02367
 M13004 V00007 V00008 V00009 V00010 V00011 V00012 V00013 V00014

VERSION V000015 V00016 V00017 V00018 V00019 V00020 V00023 V00024
KEYWORDS DNA polymerase; DNA-binding protein; RNA polymerase III;
501917.1 GI:209811
alternative splicing; coat protein; complete genome; genome-11n

SOURCE Human adenovirus type 2.
ORGANISM Human adenovirus type 2.

REFERENCE
1 (sites)
AUTHORS
Ohe, K. and Weissman, S. M.
viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovir

JOURNAL	Journal of Biological Chemistry	252 (24)	9032-9042 (1977)
MEDLINE	78046048		
REFERENCE	Doi T and Wolfsman S M		
	4 (sites)		

ADENOVIRUS
TITLE
Studies of low molecular weight RNA from cells infected with
adenovirus 2. II. Heterogeneity at the 5' end of VA-RNA I
J. Biol. Chem. 252 (24), 9043-9046 (1977)
JOURNAL
70045040
UNOTRUE

REFERENCE	REMARKS
5 (bases 10514 to 10680)	100% 0005
Pan, J., Celma, M. L. and Weissman, S. M.	
Studies of low molecular weight RNA from cells infected with	
deleting 2.111 The sequence of the promoter for VA-RNA 1	

JOURNAL
MEDLINE
REFERENCE
J. Biol. Chem. 252 (24), 9047-9054 (1977)
78046050
6 (bases 30812 to 30900)
Baker, D. C. and Roberts, D. T.

AUTHORS Zhai, B.-S. and Roberts, R. W.
TITLE Characterization and sequence analysis of a recombination site in the hybrid virus Ad2+ND
J. Mol. Biol. 120 (1), 13-31 (1978)
EDITOR

MEDLINE REFERENCE	7 (sites)
Akustajarvi, G. and Pettersson, U.	
Nucleotide sequence at the junction between the coding region and the leader sequence of the <i>hprt</i> gene and the leader sequence of the <i>hprt</i> gene	

JOURNAL
MEDLINE
79096069
8 (bases 5986 to 6236)
Proc. Natl. Acad. Sci. U.S.A. 75 (12), 5822-5826 (1978)
the adenovirus 2 hexon messenger RNA and its recombinant

AUTHORS Zilt,E.B. and Evans,K.M.
TITLE Coincidence of the promoter and capped 5' terminus of RNA for adenovirus 2 major late transcription unit
JOURNAL Cell 15 (4), 1463-1475 (1978)

MEDLINE	79084199
REFERENCE	9 (bases 21607 to 21816)
AUTHORS	Akujärvi, G. and Pettersson, U.
TITLE	Sequence analysis of adenovirus DNA. I. Nucleotide sequence

JOURNAL
MEDLINE
79119384
VIROLOGY 91 (2), 477-480 (1978)
carboxy-terminal end of the gene for adenovirus type 4 hexon
10 (bases 1 to 156; 35804 to 35937)
REFERENCE

AUTHORS Arrand, J. R. and Roberts, R. J.
TITLE The nucleotide sequences at the termini of adenovirus-2 DNA
JOURNAL J. Mol. Biol. 128 (4), 577-594 (1979)
MEDLINE 79175344

REFERENCE AUTHORS	TITLE
11 (sites) Zain,S., Sambrook,J., Roberts,R.J., Keller,W., Fried,M. and Dunn,A.R.	Nucleotide sequence analysis of the leader segments in a cl

JOURNAL	copy of adenovirus 2 fiber mRNA
MEDLINE	Cell 16 (4), 851-861 (1979), "
REFERENCE	79111249
	12 (bases 6039 to 6079: 7101 to 7172: 9634 to 9723: 10802 to 10879)

AUTHORS	TITLE
Akustjarvi, G. and Pettersson, U.	Sequence analysis of adenovirus DNA: complete nucleotide sequence of the spliced 5' noncoding region of adenovirus 2 hexon mRNA

JOURNAL Cell 16 (4), 841-850 (1979)
MEDLINE 79211248
REFERENCE 13 (sites)
AUTHORS Shinozawa, M. and Padmanabhan, R.

TITLE Nucleotide sequence at the inverted terminal repetition of

VERSION	KEYWORDS	SOURCE ORGANISM	REFERENCE AUTHORS TITLE	JOURNAL MEDLINE REFERENCE AUTHORS TITLE	JOURNAL MEDLINE REFERENCE AUTHORS TITLE	JOURNAL MEDLINE REFERENCE AUTHORS TITLE	JOURNAL MEDLINE REFERENCE AUTHORS TITLE	
J01953 J01954 J01955 J01956 J01957 K00086 K00394 K00395 K02367 M13004 V00007 V00008 V00009 V00010 V00011 V00012 V00013 V00014 V00015 V00016 V00017 V00018 V00019 V00020 V00021 V00022 V00023 V00024 V00017.1 GI:209811	DNA polymerase; DNA-binding protein; RNA polymerase III; alternative splicing; coat protein; complete genome; genome-linked protein; glycoprotein; overlapping genes; polymerase; terminal repeat; unidentified reading frame; virus-associated RNA.	Human adenovirus type 2 viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus	1 (sites) Ohe, K. and Weissman, S.M. The nucleotide sequence of a low molecular weight ribonucleic acid from cells infected with adenovirus 2	J. Biol. Chem. 246 (22), 6991-7009 (1971)	2 (sites) Jornvall, H., Ohlsson, H. and Philipson, L. An acetylated N-terminus of adenovirus type 2 hexon protein	Biochem. Biophys. Res. Commun. 56 (2), 304-310 (1974)	3 (sites) Celma, M.L., Pan, J. and Weissman, S.M. Studies of low molecular weight RNA from cells infected with adenovirus 2. I. The sequences at the 3' end of VA-RNA I	J. Biol. Chem. 252 (24), 9032-9042 (1977)
78046048			4 (sites) Celma, M.L., Pan, J. and Weissman, S.M. Studies of low molecular weight RNA from cells infected with adenovirus 2. II. Heterogeneity at the 5' end of VA-RNA I	J. Biol. Chem. 252 (24), 9043-9046 (1977)	5 (bases 10514 to 10680) Pan, J., Celma, M.L. and Weissman, S.M. Studies of low molecular weight RNA from cells infected with adenovirus 2. III. The sequence of the promoter for VA-RNA I	J. Biol. Chem. 252 (24), 9047-9054 (1977)	6 (bases 30812 to 30900) Zain, B.S. and Roberts, R.J. Characterization and sequence analysis of a recombination site in the hybrid virus Ad3+ND	J. Mol. Biol. 120 (1), 13-31 (1978)
78153757			7 (sites) Akusjarvi, G. and Pettersson, U. Nucleotide sequence at the junction between the coding region of the adenovirus 2 hexon messenger RNA and its leader sequence	Proc. Natl. Acad. Sci. U.S.A. 75 (12), 5822-5826 (1978)	8 (bases 5986 to 6236) Ziff, E.B. and Evans, R.M. Coincidence of the promoter and capped 5' terminus of RNA from the adenovirus 2 major late transcription unit	Cell 15 (4), 1463-1475 (1978)	9 (bases 21607 to 21816) Akusjarvi, G. and Pettersson, U. Sequence analysis of adenovirus DNA. I. Nucleotide sequence at the carboxy-terminal end of the gene for adenovirus type 2 hexon	Virology 91 (2), 477-480 (1978)
79119384			10 (bases 1 to 156; 35804 to 35937) Arrand, J.R. and Roberts, R.J. The nucleotide sequences at the termini of adenovirus-2 DNA	J. Mol. Biol. 128 (4), 577-594 (1979)	79175344	11 (sites) Zain, S.V., Sambrook, J., Roberts, R.J., Keller, W., Fried, M. and Dunn, A.R. Nucleotide sequence of the		

early E4 34K prote
early E4 34K prote
early E4 17K prote
early E4 33K prote
hBRF1 protein - hu
hypothetical prote
formate dehydrogen
hypothetical prote
rhodopsin - Africa
hypothetical prote
hypothetical prote
opsin, rod - Pomat
NADH dehydrogenase
E2 protein - human
NADH dehydrogenase
hypothetical prote
probable respirato
pectate lyase (EC
NADH dehydrogenase
opsin, ultraviolet
NADH dehydrogenase
NADH dehydrogenase
diphosphate (EC 3.1.5.
hypothetical prote
hypothetical prote
ALG3 protein - yea
hypothetical prote
probable membrane
hypothetical prote
heat shock protein

R; Hogenkamp, T.; Esche,

3952 H.

C;Accession: S10867; S3395
P:Hoogenkamp, T.; Esche, H.

